INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI.

17084/36

I. A. R. I. 6.
MGIPC—S4—10 AR—21-6-49—1,000.

PROCEEDINGS

OF THE

ROYAL SOCIETY OF LONDON

SERIES A—MATHEMATICAL AND PHYSICAL SCIENCES

VOL CLII

LONDON
Printed and published for the Royal Society
By Harrison & Sons, Ltd., 44-47, St. Martin's Lane
Printers in Ordinary to His Majesty

November 15, 1935

LONDON:
HARRISON AND SONS, LTD., PRINTERS IN ORDINARY TO HIS MAJESTY
ST. MARTIN'S LANE

CONTENTS

SERIES A VOL CLII

No. A 875—October 15, 1935

PAGE

Discussion on Supraconductivity and other Low Temperature Phenomena. Opening Address by J. C. McLennan, F.R.S. 1

Studies on Explosive Antimony. II—Its Structure, Electrical Conductivity, and Rate of Crystallization. By C. C. Coffin. Communicated by A. S. Eve, F.R.S. (Plates 1–2) 47

The Optical System of the Disappearing Filament Pyrometer. By F. A. Cunnold. Communicated by A. C. Egerton, F.R.S. 64

Numerical Relationships in Binary Metallic Systems. By D. Stockdale. Communicated by Sir William Pope, F.R.S. 81

Diffraction of Electrons by Metal Crystals and by Mica. By J. A. Darbyshire and E. R. Cooper. Communicated by W. L. Bragg, F.R.S. (Plates 3–4) 104

Experiments on Heavy Hydrogen. V—The Elementary Reactions of Light and Heavy Hydrogen. The Thermal Conversion of Ortho-Deuterium and the Interaction of Hydrogen and Deuterium. By A. Farkas and L. Farkas. Communicated by E. K. Rideal, F.R.S. 124

Experiments on Heavy Hydrogen. VI—The Ratio of the Magnetic Moment of the Proton to the Magnetic Moment of the Deuteron. By A. Farkas and L. Farkas. Communicated by E. K. Rideal, F.R.S. 152

On Chemical Reaction in the Electric Discharge. I—The Chemical Effects of Impulse Discharges. By E. J. B. Willey. Communicated by F. A. Freeth, F.R.S. 158

Dipole Induction and the Solvent Effect in Dipole Moment Measurements. By F. C. Frank. Communicated by N. V. Sidgwick, F.R.S. 171

A Study of Sensitized Explosions. I—The Hydrogen Oxygen Reaction Catalysed by Nitrogen Peroxide. By S. G. Foord and R. G. W. Norrish. Communicated by T. M. Lowry, F.R.S. 196

The Continuous Absorption Spectrum of Hydrogen Bromide. By C. F. Goodeve and A. W. C. Taylor. Communicated by F. G. Donnan, F.R.S. 221

The Effect of Thermal Agitation on Atomic Arrangements in Alloys—III. By E. J. Williams. Communicated by W. L. Bragg, F.R.S. 231

No. A 876—November 1, 1935

The Pressure of a Degenerate Electron Gas and Related Problems. By Sir Arthur Eddington, F.R.S. 253

The Preparation of Oxygen by Electrolysis of Baryta. By E. M. Stoddart. Communicated by F. G. Donnan, F.R.S. 273

PAGE

Aspects of the Current System Producing Magnetic Disturbance. By J. M. Stagg. *Communicated by G. C. Simpson, F.R.S.* 277

The Ultra-Violet Absorption Spectra of Some Complex Aromatic Hydrocarbons—I. By W. V. Mayneord and E. M. F. Roe. Communicated by E. L. Kennaway, F.R.S. 299

The Effect of Imprisonment of Resonance Radiation in the Decomposition of Ammonia and of Deutero-Ammonia. By H. W. Melville. Communicated by E. K. Rideal, F.R.S. 325

The Diamagnetic Susceptibilities of Salts Forming Ions with Inert Gas Configurations. II—The Alkaline Halides. By G. W. Brindley and F. E. Hoare. Communicated by R. Whiddington, F.R.S. 342

Investigations on the Combustion of Hydrocarbons. I—The Influence of Molecular Structure on Hydrocarbon Combustion. By A. R. Ubbelohde. Communicated by A. C. G. Egerton, F.R.S. 354

Investigations on the Combustion of Hydrocarbons. II—Absorption Spectra and Chemical Properties of Intermediates. By A. R. Ubbelohde. Communicated by A. C. G. Egerton, F.R.S. (Plates 5–6) 378

The Absorption of Monochromatic X-Ray Beams, of Wave-length in the Region 50 to 20 *X*-Units, in Lead, Tin, Copper, and Iron. By J. Read. Communicated by J. C. McLennan, F.R.S. (Plate 7) 402

Explosion Waves and Shock Waves. III—The Initiation of Detonation in Mixtures of Ethylene and Oxygen and of Carbon Monoxide and Oxygen. By W. Payman and H. Titman. Communicated by J. F. Thorpe, F.R.S. (Plates 8–14) 418

The Adsorption of Hydrogen on Tungsten. By J. K. Roberts. Communicated by E. K. Rideal, F.R.S. 445

Some Properties of Adsorbed Films of Oxygen on Tungsten. By J. K. Roberts. Communicated by E. K. Rideal, F.R.S. 464

Composite Films of Oxygen and Hydrogen on Tungsten. By J. K. Roberts. Communicated by E. K. Rideal, F.R.S. 477

On the Radiative Collision between Fast Charged Particles. By C. Møller. Communicated by R. H. Fowler, F.R.S. 481

No. A 877—November 15, 1935

The Energies of Nuclear Reactions. By H. A. Wilson, F.R.S. 497

On the Ground State of (H^2), the Molecular Ion (H^2+) and Wave-Mechanics. By O. W. Richardson, F.R.S. 503

Exchange of Energy between Diatomic Gas Molecular and a Solid Surface. By J. M. Jackson and A. Howarth. Communicated by D. R. Hartree, F.R.S. 515

Statistical Treatment of Strong Electrolytes. By S. Levine. Communicated by J. C. McLennan, F.R.S. 529

PAGE

The Creation of Electron Pairs by Fast Charged Particles. By H. J. Bhabha. Communicated by R. H. Fowler, F.R.S. 559

On the Instability of a Fluid when Heated from Below. By R. J. Schmidt and S. W. Milverton. Communicated by S. Chapman, F.R.S. (Plate 15) 586

Progressive Lightning—II. By B. F. J. Schonland, D. J. Malan, and H. Collens. Communicated by Sir Charles Boys, F.R.S. (Plates 16-21) 595

The Relativistic Self-Consistent Field. By B. Swirles. Communicated by D. R. Hartree, F.R.S. 625

Phase-Equilibrium of Supraconductors in a Magnetic Field. By H. London. Communicated by F. A. Lindemann, F.R.S. 650

The Nuclear Spin of Iodine. II—Fine Structure in the Arc Spectrum and a Fine Structure Perturbation Effect. By S. Tolansky. Communicated by W. L. Bragg, F.R.S. 663

The Temperature Dependence of Free Electron Susceptibility. By E. C. Stoner. Communicated by R. Whiddington, F.R.S. 672

The Interaction of Light Nuclei. II—The Binding Energies of the Nuclei H_1^3 and He_2^3. By H. S. W. Massey and C. B. O. Mohr. Communicated by R. H. Fowler, F.R.S. 693

The Diffusion of Gases through Metals. II—Diffusion of Hydrogen through Aluminium. By C. J. Smithells and C. E. Ransley. Communicated by R. H. Fowler, F.R.S. 706

The Energy of Disintegration of Radio-Phosphorus [P^{30}]. By C. D. Ellis, F.R.S., and W. J. Henderson 714

Index 725

A Discussion on Supraconductivity and other Low Temperature Phenomena

(*Held on May* 30, 1935)

OPENING ADDRESS

By Professor J. C. MCLENNAN, F.R.S.

In a discussion of low temperature phenomena it may be assumed that we are dealing chiefly with the properties of materials at temperatures that can be reached by the use of liquid helium. Until 1923 liquid helium had been produced only in Leiden. Now it is obtainable in laboratories in the United States of America, in Canada, in England, in Germany, and in the Soviet Union as well as in Holland.

NEW METHODS OF LIQUEFYING HELIUM

Since 1932 two notable advances have been made in the design of apparatus for the production of liquid helium. The equipment developed by Simon and Mendelssohn and installed in the Clarendon Laboratory at Oxford represents one of them. Liquid helium is produced at a low cost, but in very small amounts, by a discontinuous process in which the cooling is produced by the adiabatic expansion of compressed helium gas. Liquid helium has also been produced by Simon and his associates, by the desorption of helium gas from charcoal.

In the method developed by Kapitza in the Royal Society-Mond Laboratory, the gas is pre-cooled with liquid nitrogen only, and in expanding is made to do " external work " upon an expansion engine which is included in the gas cycle, and which works without lubrication at low temperatures. A temperature of 10° K is reached by means of the expansion engine, and this makes it possible to use the Joule-Thomson effect to bring about liquefaction.

PROPERTIES OF LIQUID HELIUM

For each pressure to which liquid helium may be subjected it has a maximum density at a definite and characteristic temperature. Under a pressure of 1 atmosphere, and at a temperature of 4·2° K, its density is 0·1252, or about one-eighth of that of water. Under 1 atmosphere pressure its maximum density is 0·1473, and is reached at 2·178° K. A number of the properties of liquid helium exhibit discontinuities at the temperature of maximum density. Among these are the compressibility, the specific heat, the dielectric constant, and the viscosity. Most other properties show at the same time a change of temperature coefficient.

So pronounced is the difference in the properties of liquid helium above and below 2·178° K that the two states have been designated as He I and He II, and although the transition does not seem to be accompanied by latent heat, it has been considered as in some respects a change of phase. The pressure-temperature relations, and the thermal relations of the transition from He I to He II, and of the transition from liquid to solid helium, have been extensively studied at Leiden. However, this peculiar phenomenon in the liquid helium itself is still imperfectly understood, both as regards the nature of the transition and as regards the possible structure of liquid helium II. Present indications are that He II is the stable form at 0° K and low pressures but that the transition from He II to solid helium there takes place under pressure without any energy change.

LOWEST TEMPERATURE ATTAINABLE BY BOILING UNDER LOW PRESSURE AND BY ADIABATIC DEMAGNETIZATION

The lowest temperature which has been reached with liquid helium alone is 0·71° K, but a method has recently been used for reaching very much lower temperatures. This is based on the cooling effect obtained under certain conditions by adiabatic demagnetization. Debye, and also Giauque, showed in 1926 that it would be possible to get a measurable cooling effect with paramagnetic salts if the initial temperature were sufficiently low. The method was tested by de Haas, Wiersma, and Kramers, as well as by Simon and his associates, and has been found to be highly efficient. In applying the method the salt is cooled in the presence of a magnetic field of about 30,000 gauss to a temperature of about 1·26° K by the use of liquid helium under low pressure. It is then thermally insulated, and under these conditions the magnetic field is suddenly reduced to a selected low value of a few hundred gauss or even less. Adiabatic demagnetization of the salt ensues, and as a result its temperature is lowered. The salt itself indicates by its magnetic susceptibility the temperature reached. By using this method with potassium chrome alum de Haas attained in February 1935 a temperature of 0·0044° K.

Debye has pointed out that the magnetic energy μH of a paramagnetic atom in a field of the order of 10,000 gauss is of the same order as the thermal energy kT at 1° K. Paramagnetic saturation can therefore be attained with fields of this order at liquid helium temperatures. The nuclear magnetic moments are of the order of 1/1000 of the atomic moments, and therefore saturation of the nuclear moments is to be expected for readily attainable fields in the temperature region from 0·001 to 0·01° K. It should now be possible to observe the influence of the nuclei on the magnetic susceptibility in certain cases.

SUPRACONDUCTING METALS

Up to the present time the elements given in Table I have been shown to become supraconducting. To this table should be added many alloys, including certain intermetallic compounds such as Au_2Bi, the components of which are not supraconducting at the temperatures to which they have so far been tested. One must add also certain non-metallic compounds which at ordinary temperatures are only semi-conductors.

TABLE I—SUPRACONDUCTING ELEMENTS

Element	Transition point °K
Niobium	9·2
Lead	7·2
Tantalum	4·4
Mercury	4·22
Tin	3·71
Indium	3·37
Thallium	2·37
Titanium	1·75
Thorium	1·5
Aluminium	1·14
Gallium	1·05
Zinc	0·78
Magnesium	0·70
Zirconium	0·70
Cadmium	0·54
Hafnium	0·35

It has long been a disputed question whether supraconductivity should be a general property, and whether all metallic elements would show the phenomenon at sufficiently low temperatures. The new method of attaining very low temperatures by demagnetization promises a closer approach to the answer. In particular, it would be of the greatest interest to know whether supraconductivity can be established in any of the monovalent metals or in the ferromagnetic metals.

MAGNETIC FIELDS AROUND SUPRACONDUCTORS

The series of experiments on magnetic fields around supraconductors, commenced by Meissner and his co-workers, has recently opened a new method of attack upon the problem. Four general methods have been used in the study of this problem. Meissner and his collaborators in Berlin began by studying the field distribution round cylinders, and inside a hollow cylinder, with a small test coil. Tarr, in Toronto, measured the change of flux through fixed coils round the bodies. De Haas and his collaborators in Leiden measured the penetration of the field into a cylinder by threading fine bismuth wires through the cylinder and observing the change of resistance of the bismuth. Mendelssohn and Babbitt, in Oxford, measured the magnetic moment induced in solid and hollow spheres by changes in the magnetic field, thus reviving some of the experiments performed at Leiden in the early days of the subject.

When small fields, less than the critical field required to interrupt supraconductivity, are applied to pure metals at temperatures below their transition points, the results are definite. In agreement with simple electromagnetic considerations, the supraconductor acts as if its magnetic permeability were zero, at any rate beyond a very thin surface layer. But when the transition to the supraconducting state takes place in the presence of a magnetic field, either by cooling in a steady external field or by reducing the field strength from an initial value greater than the critical field, the experiments agree only in that the result is never quite what might be expected. However, certain experimental facts seem to stand out, at least for most metals which have been tested. First, the distribution of external field is spontaneously readjusted nearly to the distribution it would have if the induction within the body were everywhere zero. Second, the field inside a hollow is only slightly altered, if at all, in the transition, showing the curious phenomenon of an apparently isolated section of magnetic field in a stable state. Third, for a solid body varying amounts of the original magnetic flux are " locked in " in the course of the transition. The field thus locked in, like the field inside a hollow, cannot then be disturbed by any external influence so long as the body remains in the supraconducting state.

Gorter and Casimir have suggested that this third phenomenon may be accounted for if it is supposed that the body passes into the supraconducting state in layers. The experiments at Toronto have shown that the amount of flux locked in for a given metal depends mainly on the shape of the sample. On the other hand, it has been shown both at Berlin and at Toronto that there is no difference in the phenomenon whether, for example, a cylinder is cooled naturally from the outside or is cooled by insulating the outside and passing liquid helium through a hole in the centre.

Along with this question of magnetic flux through supraconductors goes that of the hysteresis effects which occur when the magnetic field is increased isothermally above the critical value, and then reduced again. Other, but usually less pronounced, hysteresis effects also occur in connexion with the transition by change of temperature in a steady field.

SUPRACONDUCTIVITY IN THIN FILMS

Series of experiments have been carried out by Grayson-Smith, Misener, and Wilhelm, in Toronto, on the transition points of thin films of tin and lead. For thicknesses of the order of 10 to 1 μ there is a slight lowering of the transition point, which may be due to strains set up by unequal contraction of the film and its support. But for a thickness of tin less than 1·0 μ, and of lead less than 0·8 μ, the transition point begins to decrease very rapidly until tin film thinner than 0·3 μ could not be made supraconducting at 2° K, the lowest temperature at present attainable in Toronto. At the same time the transition range is very much widened.

The transition points of these films were found to be very sensitive to increase of the measuring current. For the thinnest film which became supraconducting, the resistance was restored by a current only one-sixth of that necessary to give the critical magnetic field at the surface. The effect of external magnetic field on the films has been studied during the past year, and it is found that when the resistance is restored in this way fields very much greater than the usual critical fields are required. This new effect is noticeable for relatively thick films, whose normal transition points do not differ appreciably from that of solid tin, up to 30 μ and greater. The transverse field required is always considerably greater than the longitudinal field. So far as it is possible to tell at present, the effects of external field and of internal current are additive when both are present, and are therefore independent. This suggests that there is a maximum current density for the supraconduction current, but that the effect has hitherto been masked by that of the magnetic field caused by the current.

Another curious result of these experiments is the effect of putting outside the tin film another coating of a non-supraconducting metal. The supraconductivity is then seriously interrupted for very much thicker films, as great as 9 μ. The free surface must have something to do with the phenomenon, although it is not necessary to its existence.

THE HALL EFFECT IN SUPRACONDUCTORS

For the investigation of the Hall effect in supraconductors Burton, Grayson-Smith, and Tarr have recently developed a completely supraconducting galvanometer, to which considerable interest is attached. The whole apparatus is immersed in liquid helium.

This galvanometer is being used to investigate not only the Hall effect but also thermoelectric effects between supraconductors. It is also being used to study the distribution of persistent currents in supraconductors.

THERMAL CONDUCTIVITY AT LOW TEMPERATURES

De Haas and Bremmer have carried out an extensive series of measurements of the thermal resistances of metals at liquid helium temperatures, both for supraconductors and non-supraconductors. For pure supraconducting metals below their transition points the thermal conductivity is increased when the electrical supraconductivity is interrupted by a magnetic field. Qualitatively this is not difficult to understand, for the electrons responsible for supraconductivity must be excluded from taking part in thermal phenomena. Quantitatively, however, there are great difficulties in reconciling these thermal conductivity measurements with other experiments.

According to theory the thermal resistance in a perfect lattice should become zero at 0° K. Actually, it reaches a minimum at some temperature in the liquid helium range, the temperature of the minimum becoming lower as the metal is made more pure. This rise of thermal resistance at very low temperatures is ascribed to the impurities and lattice irregularities, and is thus analogous to the residual electrical resistance. De Haas and Bremmer in their experiments on non-supraconductors have been attempting to disentangle the effects of irregularities from those due to the perfect lattice. This will apparently be necessary before much more can be said concerning the relations between thermal conductivity and other properties of supraconductors.

ANOMALOUS BEHAVIOUR OF CERTAIN ALLOYS

It has recently been found that certain supraconducting alloys, notably $PbTl_2$, behave very differently from the pure metals in their accompanying magnetic and thermal properties. These alloys require extremely high magnetic fields in order to restore the resistance. However, Keesom has recently found that the greatest resistanceless current which they can carry is, contrary to Silsbee's hypothesis, not nearly sufficient to give the critical magnetic field at the surface. In the work of Tarr in Toronto the flux which existed above the transition point was completely " locked in " for the alloys examined, while for all pure metals, except tantalum, all but a fraction of the initial magnetic flux was forced out when the body passed into the supraconducting state. Keesom has found that the critical magnetic field which penetrates to the centre of a cylinder at any temperature is considerably less than the critical field required to restore the resistance. In one alloy which has been measured, interruption of supraconductivity causes an irregular decrease of thermal conductivity instead of an increase. Tantalum acted more like an alloy than a pure metal in Tarr's experiments, and now Silsbee and his collaborators have found that, like the alloy studied by Keesom, tantalum is abnormally sensitive to current density.

LOW TEMPERATURE CALORIMETRY APPLIED TO RADIOACTIVE MEASUREMENTS

Reference has been made recently by Simon to the possibility of applying low temperature calorimetry to the measurement of γ-ray effects.

It would be of interest to radiologists if it should turn out to be possible to measure by this method the absorption by gaseous helium of X-rays and γ-rays of different wave-lengths. One might be able then to deduce from the results so obtained irradiation dosage in absolute units for different radiation frequencies.

APPLICATION OF THERMODYNAMICS, SPECIFIC HEATS

It has been known for some time from the work of Keesom and others at Leiden that there is a discontinuity in the specific heat of a supraconductor at the transition point. Apparently there is no measurable latent heat accompanying the transition. There is, however, an energy change of the nature of latent heat when the normal state is restored below the transition point by means of a magnetic field. A change of temperature has also been observed when the magnetic field is applied adiabatically.

The one notable success in theoretical supraconductivity has been the thermodynamical treatment by Gorter and Casimir of the connexion between these specific heat relations and the critical magnetic field. Assuming that the change of energy in a magnetic field corresponds to that for a diamagnetic body with susceptibility $-1/4\pi$ for fields up to the critical value, they have considered the thermodynamic cycle in which the body is first cooled without field, then a field applied isothermally to a value greater than the critical, and finally the body warmed again above its transition point. Assuming the condition for a reversible cycle, this gives a connexion between the specific heats and the critical field at different temperatures. The experimental values for tin and thallium are in excellent agreement with the thermodynamic calculations.

ELECTROMAGNETIC THEORY

The ordinary electromagnetic equations are untenable for supraconductors, since they become indefinite. It was supposed by Becker and his co-workers that the correct completion of the ordinary equations was to be found by taking into account the acceleration of the electrons. This gives a connexion between the rate of change of current and the electric field, so that the latter can exist momentarily while the current is changing. This treatment leads to a calculation of the current distribution, in a layer of the order of 10^{-5} to 10^{-4} cm in thickness near the surface. Meissner's experiments in which a magnetic field is spontaneously readjusted in a purely thermal transition is definitely in contradiction with the results of this method of treatment of the electrical problem. London has recently taken the view that this is evidence that the conception of accelerated electrons is wrong, and that a new form of the electromagnetic equations must be set up for supraconductors. Grayson Smith, in Toronto, has taken this as evidence for the spontaneous generation of local vortex currents, a conception which was first suggested by Frenkel. According to this conception the body would have the large diamagnetic susceptibility required to explain Meissner's experiments, and to agree with the thermodynamical treatment of Gorter. The size necessary for the smallest effective vortex agrees in order of magnitude with the critical thickness for thin films in which supraconductivity is seriously disturbed.

J. D. COCKCROFT—Professor McLennan has referred to the important contribution to the technique of helium liquefaction made during the last year by Professor Kapitza, and I wish briefly to describe this development. Most of those present will be familiar with the methods of liquefaction of helium which have hitherto been used, in particular the method in which the final cooling is produced by the Joule-Thomson effect. For this to be possible the helium must first be cooled below its inversion temperature of 40° K by the use of liquid hydrogen. Thus any laboratory wishing to use helium temperatures must necessarily acquire apparatus for the liquefaction of both hydrogen and helium.

Professor Kapitza has overcome this difficulty by applying to helium the method first used by Claude for the liquefaction of air, the method in which the gas is cooled by doing external work in an expansion engine. Most physicists must have read Claude's account of the great difficulties he experienced in finding a lubricant which would be effective at the temperature of liquid air and how a solution was reached by using liquid air itself as a lubricant. Professor Kapitza's liquefier also incorporates an expansion engine, but in his case the more serious difficulty of lubrication at helium temperatures is overcome by retaining a gap of 0·05 mm between piston and cylinder and by allowing the gas itself to act as a lubricant as it escapes past the piston, the loss being kept as low as 4% by making the piston move rapidly. The cold produced by this expansion engine cools part of the helium from the temperature of liquid nitrogen to a temperature of 10° K. This cooled and expanded helium is then used to cool the remaining part of the helium to well below its inversion temperature. This helium is then expanded and liquefied.

The liquefier takes 2 hours to produce the first litre of liquid helium and after that produces 2 litres per hour. The main disadvantage of the present design lies first in this rather long starting-time, second in the complication of the valve gear operating the inlet and exhaust valves of the helium and water. In any later models it should be possible to simplify this considerably, and there seems to be no reason why liquefiers of this type should not be built to work as efficiently and as simply as a liquid air plant so that liquid helium temperatures might become much more generally available.

The liquefier has been in frequent use during the last three months, and, provided the helium is kept pure, no difficulties are experienced in its operation. Mr. Shoenberg will describe one of the experiments carried out with its aid.

D. Shoenberg—The convenient supply of liquid helium by Kapitza's new liquefier permits the magnetic properties of supraconductors to be studied in the same way as magnetic properties at ordinary temperatures. Using a Sucksmith balance,* the force on a supraconductor in a low inhomogeneous magnetic field can easily be measured, particularly for lead, which has a relatively high critical field at the normal boiling-point of helium. The only experiment so far made was with a sphere of very pure lead (Hilger 99·999%), and the following results were observed:

(1) As the magnetic field was increased, the magnetization increased linearly corresponding to a volume susceptibility of $-3/8\pi$, the value to be expected either on classical ideas, or assuming $B = 0$.

(2) Above a certain field value the magnetization suddenly began to decrease rapidly, vanishing at about 530 gauss (close to the critical field given by Rjabinin and Schubnikow† for a long lead rod). For further increase of field the magnetization remained zero.

(3) The field at which the decrease began was very close to 2/3 of the critical field. This is just the value for which the field at the equator of the sphere becomes critical, so that supraconductivity should begin to disappear.

(4) When the field was again reduced the magnetization increased again rather *more* rapidly than the original decline and finally fell back linearly approximately along the original straight line. The possible amount of " frozen-in " lines of forces was certainly less than 10%, suggesting that the lead was purer than that used by Keeley, Mendelssohn and Moore,‡ and Rjabinin and Schubnikow (*loc. cit.*).

The magnetization curve for decreasing fields is in striking contradiction to the classical electromagnetic description of a supraconductor, and provides an independent confirmation that $B = 0$ is characteristic of the ideal supraconducting state (Meissner effect). The penetration of the field into the sphere was discussed, and it was shown that the assumption of a spheroidal boundary to the supraconducting region was too simple to account for the observed results. The small hysteresis in the magnetization curve suggested that the process was a complicated one, probably extremely sensitive to local fluctuations of field and temperature. Such detailed features of the magnetization curve require, however, further investigation.

* 'Phil. Mag.,' vol. 8, p. 158 (1929).

† 'Phys. Z. Sowjet.,' vol. 6, p. 557 (1934).

‡ 'Nature,' vol. 134, p. 773 (1934).

Professor W. H. KEESOM—*Lambda-phenomena in Liquid Helium*

Recently new measurements were made on the specific heat of liquid helium under its saturated vapour pressure.* It was ascertained that the principal change in specific heat at the lambda-point occurs within an interval of 0·002 degrees, most probably even within one of 0·0002 degrees.

Further evidence was found that there is no latent heat connected with the transformation occurring at the lambda-point.

Measurements on the specific heat at constant volume of liquid helium under pressure were made. They also show the lambda-phenomenon, as was expected from earlier measurements on the relation between pressure, volume, and temperature.

It appears to be difficult to fix a definite value for the magnitude of the jump of the specific heat at the lambda-point.

The hypothesis is made that the particular shape of the He I branch of the specific heat curve must be ascribed to the influence of fluctuations due to thermal agitation.

Viscosity and heat conductivity of liquid helium change suddenly at the lambda-point.

THE THERMODYNAMICAL TEMPERATURE SCALE BELOW 0·9° K

The thermodynamical temperature scale has been established down to 0·9° K by measurements with the helium thermometer.

For lower temperatures it can be established by experiments on adiabatic demagnetization of a paramagnetic salt followed by a measurement of the heat necessary to re-establish the initial temperature. If this heat is called Q, and the decrease in entropy connected with magnetizing the salt at the initial temperature is called S, then the temperature reached by the demagnetization experiment $T = dQ/dS$.

This method depends on an accurate knowledge of the magnetic properties of the salt at the initial temperature.†

HEAT CAPACITY OF ELECTRONS IN METALS

For silver and zinc the atomic heat at temperatures from about 6° K downwards seems to contain a part that can be ascribed to the heat capacity of free electrons according to Sommerfeld's formula, assuming one free electrom per atom.[*] Recent measurements[†] on the specific heat of KCl down to 3° K corroborate this hypothesis.

---

* W. H. Keesom and Miss A. P. Keesom, 'Physica,' vol. 2, p. 557 (1935); 'Comm. Phys, Lab. Leiden,' No. 235*d*.

† For a modification of this method so that this knowledge is not needed, *cf.* Keesom, 'Physica,' vol. 2, p. 805 (1935) ; 'Comm. Phys. Lab. Leiden,' Suppl. No. 77*c*.

The atomic heat of nickel surpasses largely the heat capacity due to the atomic lattice.[‡] The additional heat capacity can, for the range $1 \cdot 1$ to $9 \cdot 0°$ K, be represented by $C_{add} = 0 \cdot 001744\ T$. The hypothesis is made that this heat capacity is connected with the energy of the conduction electrons. The fact that it is many times larger than follows from Sommerfeld's formula for free electrons shows that in the corresponding energy band, at least at the level of the limiting energy, the density of the possible energy states is particularly large.

CALORIMETRY OF SUPRACONDUCTORS

Experiments on tin and on thallium indicated that the specific heat of the metal suffers a sudden jump when it passes from the supraconducting to the non-supraconducting state.[§] The magnitude of this jump, in absence of a magnetic field, corresponds to a formula given by Rutgers, which can be written as follows:

$$\Delta C = T \frac{d^2}{dT^2}\left(\frac{H^2}{8\pi d}\right),$$

where H = magnetic field that disturbs supraconductivity at the temperature T, and d = density.

If no external magnetic field is applied the transition from the supraconducting to the non-supraconducting state proceeds without a latent heat. If there is a magnetic field the transition is connected with a latent heat,[||] for which, assuming ideal conditions, the following equation can be derived:

$$r = -T \frac{d}{dT}\left(\frac{H^2}{8\pi d}\right).$$

---

[*] Keesom and Kok, 'Comm. Phys. Lab. Leiden,' No. 232*d*.

[†] Keesom and Clark, 'Physica,' vol. 2, p. 698 (1935); 'Comm. Phys. Lab. Leiden,' No. 238*c*.

[‡] Keesom and Clark, 'Physica,' vol. 2, p. 513 (1935); 'Comm. Phys. Lab. Leiden,' No. 235*e*.

[§] Keesom and Van den Ende, 'Comm. Phys. Lab. Leiden,' No. 219*b*; Keesom and Kok, *ibid.*, Nos. 221*e*, 230*c*.

[||] Keesom and Kok, *ibid.*, Nos. 230*e*, 232*a*.

This equation was verified for thallium in those experiments in which the metal had been brought into the supraconducting state in absence of a magnetic field.

In those experiments in which the metal had been brought to the supraconducting state a magnetic field being applied, only a certain fraction of the latent heat was measured, indicating that a certain amount of magnetic flux had been "frozen in."

The fact that the above equation is verified shows that the transition from the supraconducting to the non-supraconducting state thermodynamically is a reversible process even when persistent currents are involved. This suggests that the extinguishing of persistent currents occurring when the metal passes to the non-supraconducting state is not connected with a development of Joule-heat. The extinguishing of the persistent current directly depends on the penetration of the magnetic field.

New measurements on tin under more nearly ideal conditions are being made.

Professor W. Meissner—*The Magnetic Effects Occurring on Transition to the Supraconducting State*

In our first publication on the new magnetic effects, which take place on transition to the supraconducting state,* we have already pointed out the following: only for a solid cylinder (*i.e.*, without interior holes) of pure metal cooled in a magnetic field can one say that it behaves as if its permeability is reduced to zero when it becomes supraconducting. If one cools a hollow cylinder below its transition point in a homogeneous magnetic field, the phenomena are more complicated. This fact was confirmed by our later experiments with a cylindrical crystal of tin, having a coaxial boring.† Gorter and Casimir‡ suppose that these phenomena can be explained in the following manner: the supraconductor has, below its transition point, some non-supraconducting regions into which the lines of magnetic induction are contracted so that here the critical magnetic field strength is exceeded. In these regions the lines of induction can penetrate the material of the cylinder. In this manner Gorter and Casimir also explain how a solid sphere can retain a magnetic moment when cooled below its transition point, in spite of the new effect. The existence of this remaining magnetic moment is proved by early experiments of Onnes and his collaborators as well as by the recent experiments of Mendelssohn and Babbit.* F. and H. London,† accepting the hypothesis of Gorter and Casimir, have given a very interesting phenomenological theory of supraconductivity which contains the new magnetic effect as well as the supraconductivity itself. The important question then becomes: Is the hypothesis of Gorter and Casimir in exact correspondence with *all* of the experiments ? Does the Londons' theory then hold exactly or must it also be modified ?

---

* Meissner and Ochsenfeld, 'Naturwiss.,' vol. 21, p. 787 (1933).

† Meissner, 'Z. ges. Kälteind.,' vol. 11, p. 125 (1934).

‡ 'Physica,' vol. 1, p. 306 (1934).

This question may be considered in the light of experiments made by Dr. Heidenreich in the cryogenic laboratory of the Physikalisch-Technischen Reichsanstalt at Berlin under my direction after my transfer to Munich in April 1934. Dr. Heidenreich has studied once more the magnetic field in the internal hollow region of the tin crystal, as well as round its outer surface, and has specially remeasured the remaining field. From our earlier investigations we had already found that the lead tube as well as the tin crystal tube retain in the interior hollow regions a remaining field when the homogeneous external field is cut off after lowering the temperature beyond the transition point. Now we have found that not only the lead tube (as already mentioned in my first publication with Ochsenfeld) but also the tin crystal tube had a field remaining round its outer surface. The details, the results of which were given by diagrams, are as follows: when the temperature is lowered below the transition point, the strength of the magnetic field in the hollow of the tin crystal increases *at all points*, irrespective of whether the crystal is cooled down from the inner or from the outer surface. On the average, the magnetic field in the hollow increases by about 10%.

The direction of this field at all points has a declination of at most a few degrees from the direction of the original homogeneous field. The variations of field strength from point to point are of the order of 30%, and are not altered by taking another orientation of the crystal with respect to the original homogeneous field. The field near the outside surface of the crystal tube before taking off the outer field is not exactly that to be expected for zero-permeability. The contraction of the lines of magnetic induction is less than for a solid crystalline cylinder (without interior holes), and the variation of the field strength from point to point near the surface is more complicated. When the outer homogeneous field is taken off, there remains a field in the interior bored-out region as well as near the outer surface. When the homogeneous outer field is again established, one finds near the outer surface nearly the original field which existed before the homogeneous outer field was cut off, but in the interior space the remaining field only increases by about 3%. This increase disappears when the outer field is taken off again. On restoring and destroying the outer homogeneous field, the same change of the internal remaining field can be repeatedly reproduced. The direction of the remaining field in the interior space is the same from point to point, and it has also an inclination of a few degrees to that of the original field. The strength of the internal remaining field varies from point to point by about 20%, and is on the average 35% of that of the original homogeneous field. Near the outer surface the strength of the remaining field is about 20% of that of the original field and varies from point to point by about 100%. The variation of the direction from point to point is very complicated. One also gets the impression that at many points of the outer surface the normal component of the remaining field is different from zero.

---

* 'Nature,' vol. 133, p. 459 (1934).

† 'Proc. Roy. Soc.,' A, vol. 149, p. 71 (1935); 'Physica,' vol. 2, p. 341 (1935).

These experiments of Dr. Heidenreich confirm the statement of our first publication, namely, that the hypothesis of a vanishing magnetic permeability does not enable one to describe the results of the experiments with hollow cylindrical tubes. The recent experiments also give rise, I believe, to doubts as to whether the hypothesis of Gorter and Casimir is valid for all experiments. Especially the diagrams, taken from the observations, lead, so far as can be seen now, to the conclusion that the normal component of the remaining field at almost all points of the inner and outer surface of the crystal tube differs from zero, a fact which seems to me not to be favourable to the *general* validity of the hypothesis of Gorter and Casimir and therefore also of the theory given by the Londons.

A good check on the hypothesis of Gorter and Casimir should be the following: the critical field strength which destroys supraconductivity, being a function of temperature, the lines of induction must be contracted more and more as the temperature is lowered. Therefore the distribution of the outer remaining field lines must change on decreasing the temperature. It is remarkable that the remaining field is very much stronger for the lead tube than for the tin crystal tube. But our original idea that the remaining field, being caused by faults of the crystal lattice, may disappear entirely for an ideal crystalline cylindrical tube is not in good agreement with the fact that an orientation of the crystal tube differing from the original by about 20° did not change the orientation of the lines of induction.

Dr. R. de L. Kronig—*The Propagation of Electromagnetic Waves in Metallic Conductors and its Bearing upon the Problem of Supraconductivity*

The propagation of a plane, linearly polarized electromagnetic wave of frequency ν in a metallic conductor can be dealt with phenomenologically on the basis of Maxwell's equations if it be assumed: (*a*) that the electric current density oscillating with frequency ν at every point of the conductor has the direction of the electric field E of the wave; (*b*) that it consists of two parts, given respectively by

$$\sigma \mathrm{E} \quad \text{and} \quad \frac{\varepsilon - 1}{4\pi}\frac{\partial \mathrm{E}}{\partial t},$$

of which the first is in phase with E, the second 90° out of phase with it. σ and ε, which will be functions of ν, may be called the electrical conductivity and the dielectric constant of the metal at the frequency ν. The index of refraction n and the coefficient of extinction κ of the metal are connected with them by the equations

$$n^2 = \tfrac{1}{2}(\sqrt{\varepsilon^2 + 4\sigma^2/\nu^2} + \varepsilon),$$
$$\kappa^2 = \tfrac{1}{2}(\sqrt{\varepsilon^2 + 4\sigma^2/\nu^2} - \varepsilon).$$

If n and κ are known from experiment, σ and ε may be obtained by means of these equations.

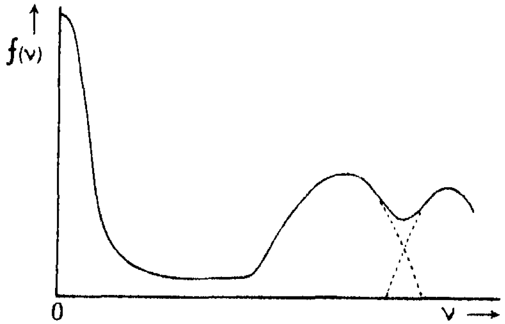
The theory of metallic conduction must furnish σ and ε in their dependence upon ν. It has been shown* that for metals in the non-supraconducting state the experimental results can be accounted for with the aid of Bloch's model of a metallic conductor, in which the electrons are supposed to move independently in the self-consistent periodic field of the crystal lattice while the electrical resistance is caused by the deviations from perfect periodicity due to the temperature agitation of the metal ions. It is then found that each volume element of the metal reacts upon the electromagnetic field as a system of harmonic oscillators in the classical theory. Denoting by $f(\nu)$ the oscillator strength per unit volume per unit frequency interval, this function is found to have the general aspect sketched in the diagram. If there were no electrical resistance for constant fields, there would be a sharp resonance frequency at $\nu = 0$, but owing to the collisions of the electrons with the lattice, caused by the temperature agitation, this is broadened into a line of finite breadth (of the order of 500 cm^{-1} at room temperature). In the region of higher frequencies (in the visible and ultra-violet) the oscillator distribution shows continuous bands, corresponding to the transitions of the electrons between the different energy zones in Bloch's model of a metallic conductor. σ and ε are connected with f by the relations

---

* R. de L. Kronig, 'Proc. Roy. Soc.,' A, vol. 124, p. 409 (1929); vol. 133, p. 255 (1931); 'Handb. der Physik,' vol. 24/2, p. 312 (1933); *see also* Fujioka, 'Z. Physik,' vol. 76, p. 537 (1932); Mott and Zener, 'Proc. Camb. Phil. Soc.,' vol. 30, p. 262 (1934).

$$\sigma(\nu) = \frac{e^2}{4m} f(\nu), \quad \varepsilon(\nu) - 1 = -\frac{e^2}{\pi m}\int_0^\infty \frac{f(\omega)\,d\omega}{\nu^2 - \omega^2},$$

e and m denoting the electronic charge and mass respectively while for the integral the principal value is to be taken.



For the alkali metals the optical properties in the visible and ultra-violet can, according to Zener,* be accounted for by assuming the bands due to electronic transitions to be exceptionally weak, so that the resonance frequency $\nu = 0$ governs the optical behaviour almost entirely. For silver, and to a lesser extent for gold, as shown by Kronig,† the broadened line $\nu = 0$ is separated by a well-pronounced gap from the first band in the ultra-violet. Here in the infra-red for wave-lengths greater than about 2μ the broadened line $\nu = 0$ is predominating, and it is then possible to make quantitative predictions regarding the optical constants.‡ Unpublished results on the reflecting power of these metals, obtained by Zernike and Lameris in the region from 1 to 13 μ at room temperature and the temperature of liquid air, and kindly put at the disposal of the author, confirm the theoretical predictions. For most other metals the continuous bands extend to much lower frequencies and merge with the broadened line $\nu = 0$, so that the theory cannot make any quantitative statements.

---

* 'Nature,' vol. 132, p. 968 (1933); *see also* R. de L. Kronig, *ibid.*, vol. 133, p. 211 (1934).

† 'Proc. Roy. Soc.,' A, vol. 124, p. 409 (1929).

‡ 'Proc. Roy. Soc.,' A, vol. 133, p. 255 (1931).

For a metal in the supraconducting state the question arises as to how the oscillator distribution will have to be modified. The experiments of de Haas and Bremmer,* showing a sudden diminution of the thermal conductivity upon passing from the non-supraconducting to the supraconducting state when reducing an outside magnetic field below the threshold value, suggest that a certain fraction α of the ordinary conducting electrons enters into a new phase, the supraconducting phase. It is plausible to identify α approximately with the relative change in thermal conductivity. According to the results of de Haas and Bremmer upon varying the temperature, α increases from the value 0 at the transition temperature the further one goes below it (*e.g.*, for tin to a value of about 35% at a temperature of 2° absolute, *i.e.*, 2° below the transition temperature). To obtain the oscillator distribution $f(\nu)$ in the supraconducting state the most plausible assumptions are that (*a*) this oscillator distribution is composed additively of two parts, due respectively to the non-supraconducting and to the supraconducting electrons; (*b*) the first part is obtained by multiplying the oscillator distribution $f(\nu)$ of the metal in the non-supraconducting state by the factor $(1-\alpha)$; (*c*) the reaction of the supraconducting electrons to an electromagnetic wave may tentatively be supposed to be that of electrons entirely free. This hypothesis is in accord both with the so-called acceleration theory, connecting current and field in the supraconducting state, as well as with the modified relation recently proposed by F. and H. London,† since for an electromagnetic wave the scalar potential vanishes so that the two viewpoints lead to identical results. Denoting by $\bar{\sigma}$ and $\bar{\varepsilon}$, σ, ε the constants of the metal in the supraconducting and non-supraconducting state respectively, we get, on the basis of the above assumptions,

$$\bar{\sigma}(\nu) = (1-\alpha)\,\sigma(\nu), \quad \bar{\varepsilon}(\nu) - 1 = (1-\alpha)\,[\varepsilon(\nu) - 1] - \alpha\,\frac{\mathrm{N}e^2}{\pi m \nu^2},$$

N signifying the number of conducting electrons per cm³. If α takes sufficiently large values, as it does when one is far enough below the transition temperature, then one should expect easily measurable changes in n and κ when passing from the non-supraconducting to the supraconducting state.

Experiments which might be suitable for a study of the effects discussed are: (*a*) the penetration of radio waves through very thin foils of supraconductors above and below the transition temperature; (*b*) the change in reflection coefficient for infra-red or visible light and the change in the state of polarization of light originally linearly polarized and reflected under an angle when passing from the non-supraconducting to the supraconducting state by reduction of an applied magnetic field below the threshold value. Another optical phenomenon which perhaps will furnish information about the supraconducting state is the fine structure of X-ray absorption edges, due to the interaction of low velocity electrons (having energies of the same order of magnitude as that of the conduction electrons) with the metallic lattice.* Changes in this fine structure might occur upon making a metal supraconducting.

---

* 'Proc. Acad. Sci. Amst.,' vol. 34, p. 325 (1931); *see also* Bremmer, 'Dissertation,' Leiden (1934).

† 'Proc. Roy. Soc.,' A, vol. 149, p. 71 (1935); 'Physica,' vol. 2, p. 341 (1935).

It should be stressed once more that the last equations for $\bar{\sigma}$ and $\bar{\varepsilon}$ are entirely tentative, but that on any other basis changes of σ and ε, and hence of n and κ, of the same order of magnitude are to be expected.

LEON BRILLOUIN, Collège de France, Paris—*Supraconductivity and the Difficulties of its Interpretation*

Supraconductive metals were first described as metals showing an extraordinary small resistivity, but when Onnes proved the most unusual duration of supra-currents it appeared that the resistivity of supraconductive metals should be actually supposed zero. The problem investigated by theoreticians was, until last year, to imagine some electronic model which could cancel all resistance inside the metal; this was hard enough, for resistance results from two different processes: first, thermic agitation disturbs the crystal lattice and thus perturbs the propagation of electronic waves (in the wave-mechanical picture); secondly, the lattice is never actually perfect but is always slightly distorted by some impurities. For ordinary metals, the thermic effect disappears at zero temperature, while the impurity effect always persists, thus preventing the resistivity from falling to naught.

Supraconductiviy arises in metals still containing some impurities, and shows a decidedly different character; so there was a first hypothesis to be introduced that supraconductive metals should crystallize in undistorted lattices, the impurities gathering in separate spots, included in holes in the perfect lattice; so the supracurrent could flow through the regular lattice, ignoring the impurities. Such an hypothesis appeared rather improbable and could never be tested; moreover, the same metals in their ordinary resistant state showed the usual impurity effect like other metals.

---

* R. de L. Kronig, 'Z. Physik,' vol. 70, p. 317 (1931); vol. 75, p. 191 (1932).

To avoid the temperature effect proved quite as difficult. I tried to imagine a very peculiar type of energy-momentum curve, which had electrons with high kinetic energies and low velocities but insensible to thermic agitation. The possibility of such unusual conditions was not actually proved: they appeared impossible in cubic or body-centred cubic lattices, but could perhaps be found in face-centred cubic or hexagonal crystals. This feature was rather satisfactory, on account of the fact that most of the supraconductors belong to these last two crystalline types.

At any rate, persistent currents could never be really stable but only metastable. This seemed inadequate to explain Onnes's persistent currents. Bloch and Landau once supposed supraconductive metals to be characterized by the fact that the most stable partition of electrons would yield a resulting current (but no mechanism whatever was suggested to explain so strange a result). Inside each small volume of the metal a certain current would flow; but in different parts of the metal the direction of the current should be different and the metal as a whole would show no current whatever, except when some external field " organized " this system of partial currents. The idea was to follow an analogy with the theory of ferromagnetism, where elementary magnetized volumes yield no total magnetization of the whole metal.

Such a model was proved impossible by a very general reasoning of Bloch himself. This remark of Bloch is of great importance and practically forbids any interpretation of supraconductivity within the frame of classical theory. Let us suppose a current I to flow through part of the metal (it might be flowing along the surface, or through some volume of the conductor); the energy of the metal will be E; we want to prove that E cannot be a minimum. If we apply a potential difference P between both ends of the conductor, then the energy of the system will be increased by a term $IPdt$, in a very short time interval dt; by changing the sign of P we can make this term positive or negative, hence we see that there is is always a possibility of decreasing the total energy E, which cannot be a minimum. Bloch's calculation is just a translation, for wave mechanics, of this very elementary result.

The only possibility, in the frame of classical theories, was to imagine some sort of current where electrons would become insensible to thermic agitation, or, in other words, the change of electronic velocities which would be produced by an external potential P (and should lower the total energy) could not be obtained by interaction of electrons and thermic agitation of the lattice. The current would not be stable but only metastable.

Metastable currents seemed practically inadequate for an explanation of older experiments on supraconductivity; and Meissner's experiment proved decisively that supra-currents were stable, thus opening the way to new theories developed by Gorter and London.

N. KÜRTI and F. SIMON—*Specific Heat of Iron Ammonium Alum below 1° K and Preliminary Determination of the Thermodynamic Scale of Temperature*

Up to the present the only information on the behaviour of the paramagnetic salts in the temperature region obtained by the magnetic cooling method was given by magnetic measurements. These measurements consisted in demagnetizing from varying initial conditions of temperature and magnetic field and determining the final temperatures in an arbitrary temperature scale (T^*) given by extrapolating Curie's law. The combination of these data with the easily obtainable knowledge of the entropy at the initial temperature as a function of the field gives us the entropy of the salt in a field zero as a function of T^*.

The addition of calorimetric measurements would permit the establishment of the thermodynamic scale, as according to the second law of thermodynamics the knowledge of both the change of heat content and of entropy between two states results in the knowledge of the absolute temperature. Moreover, the specific heat is a very sensitive indicator for the changes happening within the substance due to the interaction of the magnetic dipoles (with the lattice or with one another), as they are necessarily connected with anomalies in the specific heat.† These anomalies will occur in the neighbourhood of the temperature $\theta_m = U/R$ (U denoting the interaction energy).

For these reasons we have investigated the specific heats of iron ammonium alum, and we will report briefly on the results. As to the technique we shall only mention that the heat was supplied to the sample by γ-radiation.‡ This very convenient way of heating has the further advantage that the substance is heated absolutely homogeneously.§

Fig. 1 shows the results obtained with a cylinder of 20 mm length and 8 mm diameter. The abscissa gives the temperature T^* measured with a field of about 1 gauss parallel to the axis of the cylinder, the ordinate represents $C^* = \partial U/\partial T^*$, calculated for 1 gm-ion. The combination of this curve with the above-mentioned entropy curve gives the absolute temperature, T being the quotient of C^* over $\partial S/\partial T^*$.

---

† 'Proc. Roy. Soc.,' A, vol. 149, p. 159 (1935).

‡ 'Nature,' vol. 135, p. 763 (1935).

§ We should like to emphasize that it is not necessary to determine by absorption measurements the absolute value of the energy supplied to the sample. It is sufficient to measure relative values and to calculate one absolute value in the higher temperature region in which T^* is equal to T over a considerable range.

We find that above 0·10° K, T^* is identical with T within our experimental error of about 2%. At lower temperatures the experimental errors increase,[†] but we can state that under the conditions of our experiments beginning from this temperature the absolute temperature T is higher than the Curie temperature T^*, for instance, at $T^* = 0{\cdot}06°$ by about 10%, and we can further state that this tendency increases down to about 0·03°.[‡]

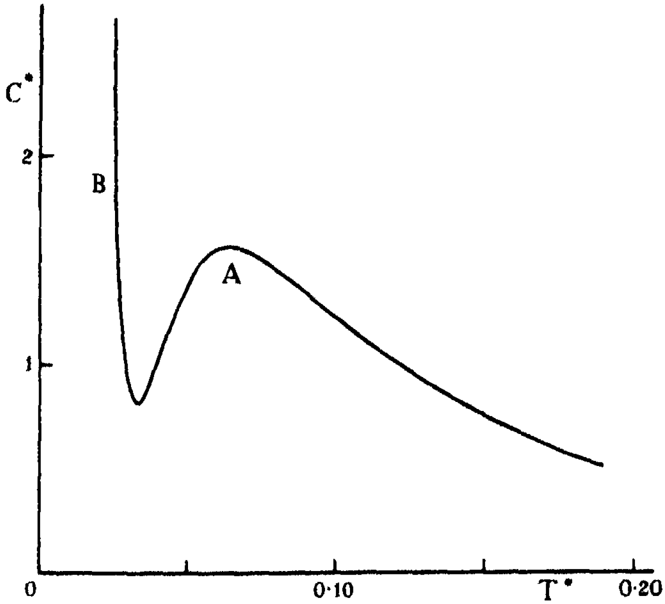


Fig. 1.

Although it is not yet possible to transform the C^*, T^* diagram of the figure to a C, T diagram with great accuracy, we can now say that the character of the curve will not be changed appreciably. Therefore we can discuss the main features of the curve, namely, the two anomalies A and B. According to Kramers, there are two reasons for the splitting up of the ground state of a magnetic ion possessing an odd number of electrons. There will be a splitting for each individual ion due to the Stark effect in the electric field of the lattice. The remaining two-fold degeneracy cannot be affected by an electric field at all and can be cancelled only by the direct action of the magnetic dipoles on each other. Thus one has to expect[†] that the anomaly in the specific heats due to the first effect will be of a flat shape (Schottky-curve), while the anomaly due to the latter effect will be of the " interaction type"; that means of the form as observed, for instance, at Curie-points. Thus we have to attribute the anomaly A to the interaction of the crystal field with the ions, and the anomaly B, of which we could previously observe only the descending part of the curve, to the interaction between the magnetic dipoles themselves. This interpretation is in accordance with experiments on diluted paramagnetic salts. With these salts not much lower temperatures are obtained[‡] when working in a temperature region down to about the maximum of the anomaly A, as the effect of the crystal field on each ion is not altered by dilution. Using higher magnetic fields, however, *i.e.*, working in the region of anomaly B, a dilution is of advantage. The interaction energy between the dipoles themselves becomes smaller, the steep fall of the entropy curve is shifted to lower temperatures, so that lower temperatures are obtained by adiabatic demagnetization than with the undiluted salt.[§]

[†] This is due to the fact that with our present electromagnet it is not possible to keep the fields very steady at high field-strengths, thus involving a rather big uncertainty for the value $\partial S/\partial T^*$.

[‡] Since this paper was read, Giauque and McDougall ('Phys. Rev.,' vol. 47, p. 885 (1935)) have published a brief account of similar experiments carried out on some gadolinium salts, and they report similar results.

Finally, we should like to emphasize that in experiments of this kind one has to pay attention to the shape of the sample, as at the high susceptibility at low temperature the demagnetizing field is of great influence. Demagnetizing, for instance, a cylinder of iron ammonium alum from 1·2° K and 16,000 gauss, an apparent temperature T^* of 0·025° is reached if the length of the cylinder is 2·5 times the diameter, whilst under the same conditions, but using a cylinder where this ratio is 1·25, the apparent temperature T^* which one observes is 0·043°.

We are repeating these experiments with an arrangement that permits a higher accuracy of the magnetic determinations and are extending them to further salts. We shall give a detailed report on the whole subject when these are completed.

Dr. R. **Peierls**—Referring to Professor Simon's paper, said that it is not surprising that metals show very small thermal conductivities at extremely low temperatures. One knows that the electric resistance of metals does not approach zero as the temperature approaches absolute zero (or the transition point in a superconductor) but that a finite resistance remains. This "residual resistance" is due to impurities and imperfection of the crystal lattice. It was first found by Grüneisen and Goens* and later derived theoretically† that the same impurities and imperfections produce an additional heat resistance which satisfies the law of Wiedemann and Franz and therefore is inversely proportional to the temperature. This resistance has to be added on to the value for the "ideal" (*i.e.*, pure and perfectly crystallized) metal which is proportional to T^2.

[†] *See* Simon, ' Ergeb. exact. Naturwiss.,' vol. 9, p. 253 (1930).

[‡] ' Proc. Roy. Soc.,' A, vol. 149, p. 172 (1935).

[§] de Haas and Wiersma, ' Physica,' vol. 2, p. 81 (1935).

Referring to Professor Keesom's paper, he said that when interpreting measurements of the specific heat of solids errors are frequently introduced by the assumption that for the part due to the vibrations of the crystal lattice Debye's law holds accurately. One easily forgets that this law is only an interpolation formula which is, in general, sufficiently accurate to describe the behaviour of the specific heat curve but not accurate enough to be applied to the determination of small deviations of the normal behaviour. Discussions of the elastic spectrum of crystals, *e.g.*, that given in recent papers by Dr. Blackman, show indeed that even in an ideal crystal there are considerable deviations from Debye's law. Certain values of the electronic specific heat can therefore only be deduced from measurements at such low temperatures that either the vibration heat is smaller than the electronic heat or that the vibrations are completely degenerate and follow a T^3-law. The first of these two conditions seems to be realized in the experiments of Professor Keesom.

F. LONDON, Oxford—*Macroscopical Interpretation of Supraconductivity*

The theory of which I should like to speak deals only with the macroscopical interpretation of supraconductivity. It seems that the principal obstacle which stands in the way of understanding this phenomenon is to be sought in its customary macroscopical interpretation as a kind of limiting case of ordinary conductivity. The present theoretical situation may be characterized in such a way that it is rigorously demonstrable that, on the basis of the recognized conceptions of the electron theory of metals, a theory of supraconductivity is impossible—provided that the phenomenon is interpreted in the usual way.

---

* 'Z. Physik,' vol. 44, p. 615 (1927).

† Peierls, 'Ann. Physik,' vol. 4, p. 121 (1930).

Bloch and Landau, particularly, have formulated the theoretical programme which seems to be indicated by the facts. They thought that one is bound to search for a model of a metal which, in its most stable states, contains a *permanent current*, and this without assistance of any external field. Often one has recalled the analogy with a ferromagnet, which in its most stable states contains a *permanent magnetization*. The thermodynamical stability of the supraconducting state, and particularly the stability of the persistent currents, seem not to admit any other conclusion. But against this conception a general theorem of the theory of electrons was quoted, even by Bloch himself, according to which the most stable state of any mechanism of electrons for general reasons may very probably have no current, and, as Professor Brillouin has just shown us; this argument can essentially be reinforced. So Bloch concluded that the only theorem about supraconductivity which can be proved is that any theory of supraconductivity is refutable, and until now experience has always verified this theorem.

We shall see that the " insoluble " problem of Landau and Bloch has really never been set by nature and that the interpretation of the facts has been in some way premature. We shall show that the phenomenon must not be interpreted as a limiting case of high conductivity, in spite of the fact that it shows electric currents, without electric fields ($J \neq 0$, $E = 0$), and that it is possible to give the phenomenon such a formulation, that it does not contradict the recognized conceptions of physics.

The experiment of Meissner and Ochsenfeld has revealed to us a new and very unexpected side of the supraconducting state. From infinite conductivity it only follows that the magnetic flux in a supraconductor must be constant, and therefore dependent upon the way in which the supraconductor passed the threshold curve. Meissner's experiment, however, has shown that in addition the magnetic flux in a supraconductor is probably exactly zero, provided the experiment is carried out under " ideal " conditions. What is to be understood by " ideal " conditions is not very easily to be defined. The fact that under " non-ideal " conditions magnetic fields have been found frozen in supraconductors seems generally to be connected with the presence of non-supraconducting inclusions, either that the substance is originally not homogeneous or that magnetic lines of force are pressed together in a part of the supraconductor in such a way that there the magnetic threshold value is exceeded.

The persistence of the original magnetic fluxes should therefore not be regarded just as the elementary phenomenon but as a very complicated matter, due to the presence of several components or phases in a complicated microscopical interpenetration. In return for that the elementary phenomenon of the pure supraconducting state should be considered as a very much simpler matter; according to Meissner's experiment, it looks as though the transition from the non-supraconducting to the supraconducting phase in a magnetic field is microscopically reversible, so far as the magnetic flux can be considered as equal to zero in any volume element of the pure supraconducting phase, independently of the way in which the threshold curve has been passed. The magnetic behaviour of a supraconductor resembles that of a very strongly diamagnetic metal, namely, of such a one which has a diamagnetic susceptibility

$$\chi = -\frac{1}{4\pi} \text{ or a permeability } \mu = 0.$$

In a diamagnetic atom we have an example of a permanent current flowing in a system which is in its most stable state. The apparent contradiction to Bloch's theorem is here avoided by the remark that Bloch's theorem deals with a system without external electric or magnetic field. We see that in a *magnetic* field this theorem evidently does not hold.

It is rather seductive to consider the supra-current as a kind of diamagnetic current, an idea which has sometimes been uttered in the past, now more seductive than ever, since Meissner's experiment seems to reveal to us the more elementary phenomenon to which one may hope to reduce the so enigmatical phenomenon of conductivity.

The macroscopical description I have developed together with H. London* shows that it is possible to work out this programme to some extent and so to escape Bloch's dilemma. The supra-current there appears as a diamagnetic current which is maintained by a magnetic field. In Meissner's experiment this field is an external magnetic field. In a permanent current in a ring the magnetic field is produced by the current itself. The most stable state of a ring has no current, unless an external field is applied. The states in which the ring possesses a permanent current are not states of lowest energy but are metastable under macroscopical conditions.

The connexion between magnetic field H and current density J we propose for the pure supraconducting phase, may be given by the equations

$$c \operatorname{rot} \Lambda \mathrm{J} = -\mathrm{H}, \qquad (1)$$

where Λ is a certain positive constant characteristic of the supraconductor.

---

*F. and H. London. 'Proc. Roy. Soc.,' A, vol. 149, p. 71 (1935); 'Physica,' vol. 2, p. 341 (1935).

We do not therefore characterize the supraconductor by a particular value of the permeability μ but by a differential equation. We shall see that this means that the total supraconductor is regarded as a single big diamagnetic atom and that the screening of an applied magnetic field is effected by volume currents instead of an atomic magnetization.

Applying the Maxwell equation

$$\operatorname{rot} \mathrm{H} = \frac{1}{c} \mathrm{J} \tag{2}$$

(we neglect here the displacement current), we can eliminate J in (1) and (2) and get

$$\Lambda c^2 \operatorname{rotrot} \mathrm{H} = - \mathrm{H},$$

or since $\operatorname{div} \mathrm{H} = 0$:

$$\Lambda c^2 \Delta \mathrm{H} = \mathrm{H}. \tag{3}$$

The solutions of this equation which behave regularly inside the supraconductor decrease exponentially as one recedes from the surface, where they are fitted into the values of the external field. In a distance of $c\sqrt{\Lambda}$ from the surface the field is practically zero. We shall see that because of the microscopical interpretation of Λ the value of this distance can be expected to be of the order of magnitude 10^{-5} cm. So we see Meissner's experiment is representable by equation (1) with one plausible restriction, namely, that the magnetic flux decreases not abruptly on the surface but continuously in a very small interval below the surface. The magnetic induction B, for reasons of simplicity, is here assumed to be everywhere equal to the magnetic field

$$\mathrm{B} = \mathrm{H}. \tag{4}$$

By taking the rotation of (3) we get an analogous equation for J:

$$\Lambda c^2 \operatorname{rotrot} \mathrm{J} + \mathrm{J} = 0. \tag{5}$$

The screening diamagnetic currents always flow very near the surface.

Differentiating (1) with respect to time and since $\dot{\mathrm{H}} = - c \operatorname{rot} \mathrm{E}$ we get

$$\operatorname{rot} (\Lambda \dot{\mathrm{J}} - \mathrm{E}) = 0,$$

or $\Lambda \dot{\mathrm{J}} - \mathrm{E}$ must be representable by a gradient of a certain scalar ϕ

$$\Lambda \dot{\mathrm{J}} - \mathrm{E} = \operatorname{grad} \phi. \tag{6}$$

For the following we need only to know that ϕ is a unique scalar. Originally we have assumed in addition that ϕ is to be identified with the electric charge times $-\Lambda c^2$. Then it would be possible to give (1) and (6) the common symmetrical form

$$\Lambda c\left(\frac{\partial J_i}{\partial x_k} - \frac{\partial J_k}{\partial x_i}\right) = f_{ik},$$

according to which the electric field would decrease to zero continuously below the boundary of a supraconductor analogously to the magnetic field. *A priori* such a small penetration of the electric field seems to be unobjectionable. But some consideration is indispensable, since these electrostatic fields can give rise to processes in which the Joule heat becomes negative, which of course is not admissible.* By the assumption of suitable boundary conditions for the electric field, however, the necessities of thermodynamics can be satisfied. On the other hand, the theory can also be formulated in a slightly different way, so that electric fields are excluded by principle. But these details about the behaviour of electric fields are not of interest to what follows and cannot be definitely decided without further experimental evidence.

Now let us consider the problem of the persistent currents or, more precisely, of the persistent magnetic fluxes. For this we need a supraconductor of the form of a ring, since a simply connected supraconductor has no appreciable magnetic flux. C may be a closed curve in the interior of the ring which cannot be deformed into a point without leaving the supraconductor. A surface S which is bounded by C lies partly in a supraconducting, partly in a non-supraconducting region. But on the whole surface Maxwell's equations hold, and therefore we get for the change of the magnetic flux through the surface S:

$$\iint_S \dot{H}_n \, d\sigma = -c \oint_C E_s \, ds.$$

The integration on the right-hand side goes only through the supraconducting region. Here because of (6) we can substitute E_s by $\Lambda \dot{J}_s - \frac{\partial \phi}{\partial s}$ and obtain:

$$\iint \dot{H}_n \, d\sigma = c \oint \left(\Lambda \dot{J}_s - \frac{\partial \phi}{\partial s}\right) ds = -c\Lambda \oint \dot{J}_s \, ds,$$

or by integration with respect to time:

$$\iint H_n(t) \, d\sigma + c\Lambda \oint J_s(t) \, ds = \iint H_n(t_0) \, d\sigma + c\Lambda \oint J_s(t_0) \, ds = \text{const.} \quad (7)$$

---

* M. v. Laue, F. and H. London, 'Z. Physik,' vol. 96 (*in the press*).

$\iint H_n \, d\sigma + c\Lambda \oint J_s \, ds$ is therefore the conservative quantity and from $\Lambda c \operatorname{rot} J = - H$ follows by applying the theorem of Stokes that the value of this conservative quantity does not depend upon the particular position of the curve C in the ring.

The ring may be so thick, and the curve C so far distant from the surface of the ring, that on C all current densities—which decrease exponentially from the surface to the interior—may be neglected. So we obtain

$$\iint H_n (t) \, d\sigma = \iint H_n (t_0) \, d\sigma. \qquad (7')$$

A magnetic flow which is once present in the hollow of the ring remains constant for any time, whatever perturbing external fields may be applied. The ring behaves like a permanent magnet. But, as it contains the magnetic energy of the flux, the ring is to be regarded as being in a *metastable* state. Only by a *finite* variation of the parameters of the system (*i.e.*, by transgressing the threshold value) can it change into the absolutely stable state which contains no flux.

It may be emphasized that our purely magnetic description is thoroughly different from a conception which has sometimes been uttered according to which the supraconductivity should be characterized by the particular value $\mu = 0$ of the magnetic permeability. For simply connected supraconductors, it is true, both formulations give macroscopically identical results for the merely magnetic phenomena. But for multiply connected supraconductors the two formulations are thoroughly different from each other. A ring which is characterized by a permeability has in its hollow a field, which is proportional to the field-strength of the applied external magnetic field and which in particular is necessarily zero, if no external field is applied. From $\mu = 0$ only, the persistence of the ring currents or of the magnetic flux through the hollow cannot be derived. On the other hand, the diamagnetic currents of a metal of a certain permeability have no component normal to the surface of separation, where it borders upon another metal. The $\mu = 0$ formulation therefore cannot possibly represent the fact of the possibility of tapping a current out of the supraconductor.

In our equation $\Lambda c \operatorname{rot} J = - H$ the magnetic field in a supraconductor is represented in the same form by the current density J as by a vector potential A. For the current the absolute value of its components has a physical meaning; the vector potential A, however, can always be replaced by an equivalent A′ if

$$A' = A + \operatorname{grad} k,$$

where k is an arbitrary single valued function of space. If Maxwell's condition div $A = 0$ is postulated to be also valid for the A' we have to impose on k the condition

$$\Delta k = 0.$$

In a *simply* connected supraconductor k always can be determined so that not only the rotations of A and J but also A and J themselves become proportional to each other:

$$A = -\Lambda c J. \tag{8}$$

In *multiply* connected supraconductors, however, this is not generally possible. If there is a magnetic flux F through a hollow we have

$$\oint A_s\, ds = \iint H_n\, d\sigma = F \neq 0,$$

if one integrates round the hollow, while

$$\oint J_s\, ds = 0,$$

provided that the path of integration lies sufficiently deep in the interior of the supraconductor. Therefore $\Lambda c J$ and $-A$, having the same rotation, must differ from each other by the gradient of a multiply valued scalar ν, the increment of which is equal to the magnetic flux through the hollow:

$$\Lambda c J + A = \operatorname{grad} \nu. \tag{9}$$

Also in this case we can give a distinguished choice of A, which is a consequent generalization of our statement $A + \Lambda c J = 0$ for simply connected supraconductors. In addition to $\operatorname{rot}(A + \Lambda c J) = 0$ and $\operatorname{div}(A + \Lambda c J) = 0$ in the supraconductor we postulate

$$A_n + \Lambda c J_n = 0$$

at the surface. That gives for ν the following conditions:

$$\Delta \nu = 0 \quad \text{in the supraconductor}$$
$$\frac{\partial \nu}{\partial n} = 0 \quad \text{at its surface.}$$

By the condition that the increment of ν in a circuit around the ith hollow must be equal to the magnetic flux F_i through that hollow—or, to speak more exactly, equal to the value of the conservative quantity $\iint H_n\, d\sigma + \oint \Lambda c J_s\, ds$ of this hollow—ν is uniquely determined and is found to be independent of time.

The equations (8) and (9) are equivalalent to the fundamental equation (1).

The foundation of our macroscopical equations by the theory of electrons in metals has not yet been undertaken. But it is rather attractive to try to sketch the programme which seems to be set by our equations to a future microscopical analysis.

The density of the electric current is known to be given in quantum mechanics by

$$J = \frac{he}{4\pi im}(\psi \operatorname{grad} \psi^* - \psi^* \operatorname{grad} \psi) - \frac{e^2}{mc}\psi\psi^* . A. \qquad (10)$$

Let ψ be the wave function of a single electron in the self-consistent field of the others, then $\psi\psi^*$ gives the value of the statistical expectation for this electron at every point of the space. Summing over all electrons $\Sigma\psi\psi^*$ gives the number n of electrons per cm^3.

If one were to use the original eigenfunctions of the free electrons in the metal, the terms in brackets, for reasons of symmetry, would vanish in summing over all electrons and equations (10) would become identical with (8), the constant Λ being equal to m/ne^2.[†] But, of course, actually the eigenfunctions of the electrons in the metals are very considerably disturbed by a magnetic field and therefore the terms in brackets in (10) do not vanish. On the contrary, they are of the same order of magnitude as the terms containing the vector potential. Moreover, the energy values of the electronic states become disturbed and their thermal distribution therefore changes. The result is only a very weak diamagnetism; the so-called Landau-Peierls diamagnetism.

But suppose the electrons to be coupled by some form of interaction, in such a way that the lowest state may be separated by a finite interval from the excited ones. Then the disturbing influence of the field on the eigenfunctions can only be considerable if it is of the same order of magnitude as the coupling forces. So long as the magnetic field is sufficiently weak, we may assume the perturbation of the eigenfunctions to be proportional to the square or higher powers of the magnetic field. We write it in the form:

$$\psi = \psi_0 + H^2\psi_1,$$

where ψ_0 is the eigenfunction without field. With this eigenfunction the bracket in (10) gives only a contribution which is quadratic in H and can be neglected in comparison with the last term, which contains A. Such a "rigid" behaviour of the eigenfunction, therefore, would be able to represent the relation (8) between magnetic field and current we have proposed.

[†] Here we leave the question of the exact meaning of n open.

Incidentally, this mechanism is in no way an absolutely new one. It characterizes the whole supraconductor as a single big diamagnetic atom. In a magnetic field the eigenfunction of a diamagnetic atom undergoes a perturbation proportional to H^2 only, which is of no account for the diamagnetism. The diamagnetic current is exhaustively represented by the term $e^2/mc\,\psi\psi^* A$ of (10), where ψ is the non-perturbed eigenfunction, the moment of that current gives immediately the well-known formula for the diamagnetic susceptibility of the atom. The ordinary atoms are so small that an external magnetic field is only very little screened from their interior. Only an atom of a diameter of $\gtrsim 10^{-5}$ cm (with a density of electrons 10^{23} cm^{-3}) would prevent practically completely a magnetic field from penetrating into its interior.

A supplement is still wanted for a supraconductor of the form of a ring. For such a one we have the equation (9) instead of (8):

$$\Lambda c J + A = \text{grad}\ \nu, \tag{9}$$

where the term grad ν is necessary in order to describe the magnetic flux through the ring.

Let us consider an atom of the shape of a ring, having a magnetic flux F through its hollow, which at first may be confined to such a small region of the hollow that nowhere does it touch the ring. It is inevitable that the vector potential A of such a magnetic field must be essentially different from zero, even in the ring, where in reality no magnetic field is present. For in every closed curve which surrounds the magnetic flow, in particular on such a one in the ring, the equation $\oint A_s\,ds = F$ must hold. Therefore in the wave equation of the ring electrons we have inevitably a vector potential, although in reality they are not in a magnetic field. One compensates the presence of such unphysical vector potentials —which can always be written in the form of a gradient of a multiply valued scalar—by " gauging " the eigenfunction. Here A = grad ν and the correctly gauged eigenfunction is to be written

$$\psi = \psi_0 e^{-\frac{2\pi i}{h}\frac{e}{c}\nu},$$

here ψ_0 is the eigenfunction of the ring-electron without flux. It is immediately verifiable that this ψ is an exact solution of the wave equation with the unphysical vector potential A = grad ν and that the current (10) of this ψ vanishes if ψ_0 had no current.

Now suppose that the magnetic flux slightly penetrates into the ring; but it may be sufficiently weak, so that the magnetic field is always smaller than the threshold value. Then the vector potential may be written in the form

$$\mathrm{A} = \mathrm{grad}\ \nu + \mathrm{A}_1.$$

Now the part A_1 is only a small supplement and may perturb the eigenfunction as little as we have assumed generally to be the case in weak magnetic fields. Then only this supplement A_1 gives rise to a current in (10)

$$\mathrm{J} = -\frac{e^2}{mc}\,\Sigma\psi\psi^*\,\mathrm{A}_1 = \frac{1}{\Lambda c}\,(\mathrm{grad}\ \nu - \mathrm{A}).$$

This is exactly our fundamental relation (9) for the ring. The persistent currents in a ring are therefore really to be conceived as a diamagnetic phenomenon which stabilizes itself.

A future foundation of this theory by means of the theory of electrons in metals has therefore apparently to explain a relatively simple state of affairs, namely, that when a supra-current is generated or maintained practically *nothing* happens, at least as concerns the behaviour of the wave function of the supraconducting state. Or, in order to say it more exactly, it would be sufficient to show that in a magnetic field the wave function of this state undergoes no disturbance proportional to H but only such a one proportional to a higher power of H. It is characteristic of the wave description of matter that the waves do not represent immediately the state of motion of a particle in the presence of magnetic fields. The wave number of the De Broglie wave determines the total momentum p, which in a magnetic field is composed of the " kinetic " momentum mv, plus the "*potential*" momentum $e/c\,\mathrm{A}$. If the total momentum of the degenerate gas of electrons is equal to zero and remains zero even in a magnetic field in consequence of interactions which are still to be explained, the current is proportional to the negative potential momentum and our equations hold.

Professor W. MEISSNER—The theory of the Londons has the result that there may be an electrostatic field in a supraconductor, a result that practically never may be verified. It would be of great interest to hear from Dr. London whether this difficulty can be avoided by modifying his theory.

Dr. LONDON (*in answer to Professor Meissner*)—I am not aware that anyone has ever seriously tested how far electrostatic fields penetrate into a supraconductor. The continuous decrease of the electric field in a minute layer below the surface of the supraconductor, as assumed in the original simple form of our theory, would be appropriate for describing our, at present rather rough, experience concerning the breakdown of the electric voltage in a supraconductor. Thus we cannot find in the facts any grounds for excluding these minute electric fields by principle, although their presence seems to be somewhat strange. On the other hand, formulations without electrostatic fields are quite possible in the framework of our theory, and two of them have already been discussed some time ago.* In one of them the potential fields are excluded from the very beginning, in the other one the presence of normal conducting electrons effects the complete breakdown of the field. It will perhaps be possible to decide upon these two possibilities by experiment.† Nevertheless, the question of the non-existence of electrostatic fields cannot yet be regarded as definitely settled. If they should exist a condenser, when its plates become supraconducting, would show a small change of its capacity. An experimental test of this question is being undertaken at present in the Clarendon Laboratory.

K. MENDELSSOHN—The experiments carried out in Oxford were mostly concerned with the induction and energy content in various supraconductive substances. Meissner and Ochsenfeld‡ had observed that the induction in tin cylinders which were cooled in a magnetic field became zero at the threshold value of supraconductivity, and Gorter§ based a thermodynamical treatment on the assumption that zero induction distinguishes the supraconductive from the normal state. It must be emphasized, however, that the coincidence of zero resistance (R) and zero induction (B) has to be treated as a purely empirical fact which cannot be derived, in general, by electrodynamical considerations. It seemed therefore desirable to test whether for all supraconductive substances ($B = 0$) is always characteristic for the supraconductive state and whether the caloric effects which can be expected according to Gorter's treatment actually occur.

THE CHANGE OF INDUCTION

Fig. 1 gives a typical threshold curve of a supraconductor, which, if ($B = 0$), indicates the difference of free energy between the normal and the supraconductive state.

---

* 'Physica,' vol. 2, p. 341 (1935), § 6, § 7.

† 'Nature,' vol. 133, p. 497 (1934).

‡ 'Naturwiss.,' vol. 21, p. 787 (1933).

§ 'Nature,' vol. 132, p. 931 (1933), and Casimir, vol. 1, p. 306 (1934).

At first the change of induction in solid and hollow tin spheres was determined by two different methods.* With each method observations were carried out by passing the threshold curve at constant field by varying the temperature ($p_4 \rightarrow p_3 \rightarrow p_4$) and at constant temperature by varying the field ($p_2 \rightarrow p_5 \rightarrow p_2$). It was found in all cases that when passing the threshold curve from the normal to the supraconductive state in any point but T_0 the induction decreased but did not change to zero. For instance, when we had cooled the spheres in zero field ($p_1 \rightarrow p_2$), then established a field at constant temperature ($p_2 \rightarrow p_3$), the induction of the specimen was zero as long as the threshold value was not surpassed, as can be predicted in this special case from (R = 0) by pure electrodynamics. When further increasing the field ($p_2 \rightarrow p_5$) the spheres passed into the normal state and the induction in them became normal. On decreasing the field, however, the induction did not change to zero again, but part of the magnetic flux was "frozen in," and this amount was considerably larger in the hollow than in the solid sphere. As these experiments indicated that the reversibility of the change in induction depended on the geometrical form of the specimen in all following experiments, only long-shaped specimens in a longitudinal field were employed.

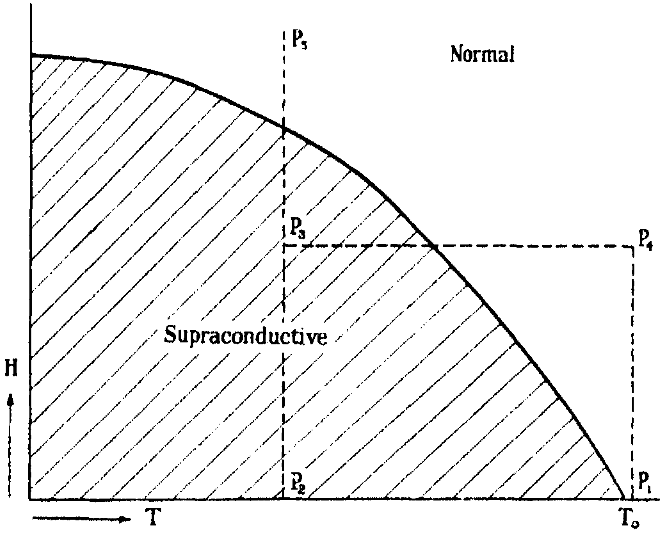


FIG. 1.

The supraconductors for which the change of induction was investigated are: mercury, tin, lead, tantalum, and various alloys (SnBi, SnCd, PbBi, $PbTl_2$).* Mercury was the only substance in which the change was entirely reversible. In tin single-crystals, polycrystalline tin and lead, the " frozen in " flux amounted from 6% to 15% of the total flux threading the specimen at the threshold value ($p_2 \rightarrow p_5 \rightarrow p_2$); the percentage in tantalum is considerably higher and varies for different samples. In all alloys the entire flux which penetrated the specimen " froze in." The penetration by lines of force, however, occurred in several alloys already far below the threshold curve.† As these experiments seemed to indicate that the amount of " frozen in " flux depended mainly on the purity, lead with 1%, 4%, 10% bismuth was investigated, and the results actually showed that the " freezing in " increased with the addition of the second component.

---

* Mendelssohn and Babbitt, 'Nature,' vol. 133, p. 459 (1934).

This dependence of the percentage of the residual flux on the purity of the substance seems to have a deeper significance when we remember that alloys usually have much higher threshold values than pure supraconductors. That means, however, that those parts of the specimens where the impurity or a purposely added second component forms an alloy will easier become supraconductive than the rest of the specimen. It would be " rings," formed by these regions of higher threshold value, which are responsible for the " freezing in." In alloys which contain an appreciable amount of a second component such " rings " will form a " sponge " of high threshold value throughout the material, which causes all lines of force penetrating the specimen to be caught by persistent currents. As in many metals impurities are to be found mainly on the boundary surfaces of the crystallites, it seems quite possible that the number of " frozen in " lines of force depends on the size of the crystals, and whether in the preparation of the specimen a method was employed which promoted such a " self-purification " as, *e.g.*, very slow growth of the crystal.

Further, we know that the threshold value of a substance is changed by a deformation of the crystal lattice as, *e.g.*, external pressure,‡ so that not only impurities but in general all disturbances of the lattice may, if sufficiently numerous, create a " sponge " of higher threshold value in the supraconductor.

Experiments were further carried out to determine the width of the transition between the supraconductive and the normal state.

---

* Keeley and Mendelssohn, ' Nature,' vol. 134, p. 773 (1934).

† *See also* de Haas and Casimir-Jonker, ' Nature,' vol. 135, p. 30 (1935).

‡ Sizoo and Kamerlingh Onnes, ' Comm. Phys. Lab. Leiden,' 180*b*.

The results obtained so far* seem to show that the width of the transition and the coincidence of the decrease in induction and resistance in pure metals depends very much on size, shape, and arrangement of the crystals and on the purity of the material. Whether the change in induction in extremely pure single crystals will be entirely discontinuous is, however, still an open question, the answer to which can only be given by more accurate experiments.

We have seen that the magnetic irreversibility increased with growing contamination of the metal and, finally, the transition became completely irreversible in alloys. In the same way a gradual increase in the width of the transition region can be observed. Whereas in impure substances the transition region is a mere broadening of the threshold curve, it plays an overwhelming rôle in alloys.

As mentioned above, we observed in our experiments that at all temperatures below the normal transition point (r_0) lines of force entered alloys which previously were cooled in zero magnetic field at tangential field strength which were far below the corresponding threshold values of supraconductivity. This is in agreement with our assumption that a supraconductive alloy is constituted like a " sponge " of high threshold value. The meshes of such a " sponge " are unable to keep out the magnetic field when the material enclosed in them has passed into the normal state, because the current induced in the individual meshes or rings will soon overcome the threshold value.

This assumption has quite recently been tested and proved by experiments on lead-bismuth,† where it was found that the meshes could be broken up and " filled " with magnetic flux by increasing the field, but kept the " frozen in " flux when the field was decreased for the same amount.

When carrying out magnetic experiments on alloys, it therefore has to be remembered that as soon as a " sponge " is formed no information on changes of induction inside the meshes of the " sponge " can be obtained unless the current threshold value in them is exceeded. That means that considerable changes of induction in parts of the alloy may take place which only result in an inhomogeniety of the field inside the meshes but in no change of the " gross " induction of the specimen which is determined in the experiment. Further evidence on these questions, however, can be obtained by calorimetric experiments.

---

* Unpublished measurements with Dr. A. Meetham.

† Mendelssohn and Moore, ' Nature,' vol. 135, p. 836 (1935).

HYSTERESIS

The experiments on the magnetic reversibility of the transition between the supraconductive and the normal state have shown that marked magnetic hysteresis phenomena can occur when lines of force are " frozen in," *i.e.*, when at the transition from the normal to the supraconductive state non-supraconductive enclosures have been left in the specimen. The question arises whether this magnetic hysteresis is connected with, and if so whether it is the only cause for, the hysteresis phenomena which occur in conductivity experiments. Certainly the supraconductive cross-section of a wire and current distribution in it will be different, whether the wire contains non-supraconductive enclosures or not, and some observations* seem to indicate that conductivity hysteresis only appeared when parts of the conductor had a shape, which distorted the field and probably caused " frozen in " lines of force.

On the other hand, it is evident that conductivity hysteresis mainly occurs in very pure substances which seem to " freeze in " only a small fraction of the flux they contained in the normal state. Some features in our experiments on the change of induction let it appear possible that another kind of hysteresis may be expected at the transition between the normal and the supraconductive state. We have seen that the width of the transition region varies from an almost discontinuous jump for pure substances to very broad regions in the H, T-diagram for alloys. With decreasing temperature or field, supraconductivity will occur first in the regions of high threshold value in an alloy or impure metal and these regions will form, therefore, nuclei for the further spreading of the supraconductive state in the specimen. In a very pure metal (single crystal) no such nuclei exist owing to the homogeneity of the material, and it is conceivable that phenomena resembling the supercooling of a very pure liquid below the melting-point could occur. Such a delayed transition from the normal to the supraconductive state would show phenomena as conductivity hysteresis. We hope the magnetic experiments will give evidence whether this supercooling exists.†

CALORIMETRIC EXPERIMENTS

Keesom‡, Rutgers,§ and Gorter|| have shown that the thermal and magnetic phenomena in supraconductors are closely related and that the

---

* de Haas, ' Leipziger Vorträge ' (1932).

† Since then experiments on pure mercury have given clear evidence of such delayed magnetic transitions.

‡ ' Rapp. 4e Com. Phys. Solvay,' p. 288 (1924).

§ ' Physica,' vol. 1, p. 1055 (1934).

|| *Loc. cit.*

difference in specific heat between the supraconductive and the normal state and the magnitude of the heat of transition could be predicted quantitatively if $(B = 0)$ was assumed for the supraconductive state.

We first observed the heat of transition between the normal and supraconductive state in adiabatic experiments on tin.* A very accurate isothermal determination on thallium was carried out by Keesom and Kok.† These experiments have shown that only a slight irreversibility occurred in a pure metal, thus corroborating the results obtained in the magnetic investigations.

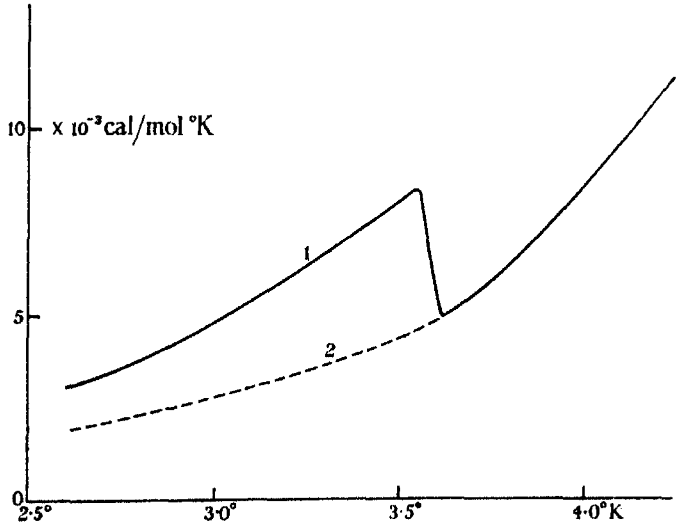


FIG. 2—Gives the preliminary result of some calorimetric experiments since carried out by Miss J. R. Moore, which seem to give direct evidence of our "sponge"-model. The specific heat of tin which was alloyed with only 4% Bi has been determined with and without "frozen in" field. Curve 1 indicates the specific heat when the specimen was cooled in zero magnetic field. Curve 2 was obtained by switching on and off a high field at the lowest temperature. It is evident that supraconductivity inside the meshes of the "sponge" has been suppressed by the "frozen in" field.

At the normal transition point of a supraconductive alloy an anomaly in the specific heat should exist if the alloy behaved in the same way as a pure metal (Sn, Tl), *i.e.*, if zero induction and zero resistance always coincided. This discontinuity could be much greater as in pure metals on account of the far steeper threshold curve of some alloys. We therefore determined the specific heat of $PbTl_2$,[*] of which the threshold values were known to be high. The measurements, however, showed no sign of a discontinuity of such order.

---

* Mendelssohn and Moore, 'Nature,' vol. 133, p. 413 (1934).

† 'Physica,' vol. 1, p. 503 (1934).

This result fits with the observation that in the greater part of the supraconductive region of the H, T-diagram the induction is not zero. It further supports our assumption of a " sponge " of high threshold value and indicates that the skeleton of the sponge can take up only a relatively small volume if it is distinguished by zero induction.

We have emphasized, however, in a previous note that the description of the supraconductive " state " by zero induction loses its significance as soon as supraconductive regions of atomic dimensions or only of such a size are considered which is comparable with the depth to which a magnetic field enters a supraconductor on the boundary surface.[†] This depth has been calculated by Becker, Heller, and Sauter,[‡] and by F. and H. London,[§] and was found to depend on the number of supraconducting electrons. Assuming that their number is of the order of the number of atoms, this depth amounts to about 10^{-6} cm. The relative volume then taken up by the sponge can be greater if the skeleton is sufficiently fine.

In respect to the thermal behaviour of a supraconductive alloy, however, both models will result in the same effects.

Summing up the results of these experiments it seems to us that (within the accuracy of the methods used) Gorter's ideal model of a supraconductor of zero induction is realized in pure, undistorted single crystals. It must be left to further experiments to show whether the change of induction in an external magnetic field is actually discontinuous or extended over a temperature interval, and whether, therefore, the heat liberated at the transition is a " latent " heat that occurs at one definite temperature or is an anomaly of the specific heat, which never becomes infinite. A decision of these questions will not alter Gorter's thermodynamical formulae considerably, but it will be of great importance for the conception of a supraconductive and a normal " phase." If the transition is not discontinuous, *i.e.*, if an intermediate state exists which covers a wider region in the H, T-diagram as has to be ascribed to the statistical deviations, " phase "-boundaries, for instance, would not exist. The existence of such " phase "-boundaries in " ideal " supraconductors has not yet been investigated.

---

[*] Mendelssohn and Moore, ' Nature,' vol. 134, p. 773 (1934); ' Proc. Roy. Soc.,' A, vol. 151, p. 334 (1935).

[†] The behaviour of small supraconductive regions has recently been treated more theoretically by Gorter (' Physica,' vol. 2, p. 449 (1935)).

[‡] ' Z. Physik,' vol. 85, p. 772 (1933).

[§] ' Proc. Roy. Soc.,' A, vol. 149, p. 71 (1935).

The interpretation of the phenomena by the assumption of two " phases " seems less significant in alloys than in pure metals. The " phase "-assumption will be justified if the regions of high threshold value, the skeleton of the " sponge," form a supraconductor similar to a pure metal, for which the threshold curve indicates the difference of free energy between the supraconductive and normal " phase." It should be possible under these conditions to produce macroscopic supraconductive regions of high threshold value, of which the specific heat would follow Rutgers's formula and the magnetic behaviour would be similar to that of pure mercury. Although the exclusion of this possibility would be premature, it needs emphasizing that no indication of such a behaviour has been found so far in the experiments.

It will be impossible in principle, however, to produce macroscopic regions of such a " high threshold value supraconductor," and the description by two " phases " will lose all significance if supraconductivity under high fields is confined to regions which are of the order of the penetration depth of the magnetic field or perhaps even atomic chains. That means: the skeleton of the sponge would be very fine but not necessarily have zero induction and could occupy a considerable fraction of the entire volume of the alloy. Finally, it is possible that the number of supraconductive electrons in an alloy of high threshold value is small and magnetic penetration accordingly greater than 10^{-6} cm. In such a case the supraconductive regions of high threshold value and ($B \neq 0$) could be larger in size without changing the thermal properties of the alloy. In all these latter cases the general characteristics for the distinction of two phases, the " phase " boundary and the difference in free energy defined by the threshold curve, will vanish.

We think that all experimental results so far obtained on impure metals and on alloys can be explained by their inhomogeneity which causes the formation of a " sponge " of higher threshold value. Such a model would substitute the *ad hoc* assumption of a fine partition in supraconductive and normal regions in Gorter's recent theoretical treatment and thereby avoid a new assumption for which otherwise no simple reason could be given. Our explanation seems the more plausible as we are able to trace the gradual change of the phenomena with increasing inhomogeneity and impurity, and as some experiments seem to indicate directly the existence of such a " sponge."

Mr. J. D. **Bernal** remarked on the relationship between crystal chemistry and physics and super-conductivity.

Metals showing super-conductivity may be divided into two groups. One group has high characteristic temperatures and is paramagnetic, while the other group has low characteristic temperatures and is diamagnetic, with certain exceptions. It seems probable that superconductivity in the two groups is of a different physical origin, although of an electronic nature in both cases.

This is strongly supported by the occurrence of super-conductivity among compounds. Among the B group those compounds, and only those compounds, are super-conducting for which the mean number of conduction electrons per atom lies between 2 and 4·2. In the A group no such simple rule holds, but interstitial compounds of non-super-conducting metals such as Mo_2C and FeB are themselves super-conducting.

Crystal structure, however, has apparently no relation to superconductivity. Every variety of metallic crystal structure is found among super-conductivity compounds. Again, as regards super-conductivity, there is no sharp distinction between a pure compound and a solid solution.

N. F. **Mott**—Professor Keesom has reported that at low temperatures the atomic heat of nickel surpasses the heat capacity due to the lattice by a term which, in the range 1·1–9·0° K, can be represented by

$$C_{\text{add}} = 0 \cdot 001744 \text{ T cal/gm atom.}$$

I wish to point out the connexion between this term and the excess specific heat of this element at high temperatures. As is well known, the specific heat of nickel rises to a sharp maximum at the Curie temperature, 358° C, and this is satisfactorily explained, in terms of the Weiss theory, as representing the work required to destroy the spontaneous magnetization. Even above the Curie temperature, however, where no spontaneous magnetization remains, the atomic heat is considerably greater than for other metals, and increases with increasing temperature. To show this, we give the *minimum* values $(C_p)_{\min}$ of the atomic heat at any temperature above the Curie point obtained by various investigators:—

Investigator	$(C_p)_{\min}$	Reference
Sucksmith and Potter ..	7·9	'Proc. Roy. Soc.,' A, vol. 112, p. 157 (1926).
Lapp	7·4	'Ann. Physique,' vol. 12, p. 442 (1929).
Klinkhardt	7·3	'Ann. Physik,' vol. 84, p. 167 (1927).
Ahrens	7·8	'Ann. Physik,' vol. 21, p. 169 (1934).
Grew	7·6	'Proc. Roy. Soc.,' A, vol. 145, p. 509 (1934).

The atomic heat of copper at a corresponding temperature, 400° C, is 6·53.

The only measurements for nickel which extend to temperatures much above the Curie temperature are those of Klinkhardt, who finds that at 800° C_p is 8·29.

The above values refer to the atomic heat at constant pressure; to obtain the atomic heat at constant volume one must subtract the dilatation correction

$$C_p - C_v = \frac{T}{J}\frac{\alpha^2}{\chi}V,$$

where

V = atomic volume

α = thermal expansion coefficient

χ = compressibility

J = mechanical equivalent of heat

T = absolute temperature.

At 400° C this correction amounts to 0·3 for nickel and to 0·4 for copper.* We see, therefore, that above the Curie point the atomic heat of nickel exceeds the classical value 3R = 5·97 by about 1·2–1·5 cal/gm atom. This term has been called, by Lapp and others, the " unknown term " in the specific heat.

Born and Brody† have shown that the heat capacity of the lattice should exceed the value 3R at high temperatures owing to the anharmonic terms in the expression for the energy of a lattice wave; but it is improbable that the additional term would be much greater for nickel than for copper and for similar metals; we may therefore conclude with confidence that the additional specific heat is due to *electrons.*

In a recent paper,‡ I have discussed some of the observed properties of nickel and of its alloys in the light of the electron theory of metals, and have deduced the following results. In the metallic lattice there are two bands of allowed energy levels, corresponding to the 3*d* and 4*s* states of the free atom. In the 4*s* band 0·6 electronic states per atom are occupied, and in the 3*d* band there are an equal number of unoccupied states (" positive holes "). The latter are responsible for the ferromagnetism, the number 0·6 being actually obtained from the intensity of the spontaneous magnetization at low temperatures. The 4*s* electrons have energies lying in a range of about 6 electron volts; they are responsible for the cohesive forces in the metal, and, up to the melting-point, their state is almost independent of temperature.

---

* *Cf.* Euken, ' Handb. Exp. Physik,' p. 211.

† ' Z. Physik,' vol. 6, p. 132 (1921).

‡ ' Proc. Phys. Soc.,' vol. 47, p. 571, 1935.

According to recent calculations * for sodium and for copper, the *s* electrons in the lattice behave as though they were free; the field of force in the lattice may therefore be neglected in calculating their specific heat; consequently the electrons will make a contribution to the heat capacity *per electron* given by the Sommerfeld formula

$$c_v = \tfrac{1}{2}\pi^2 k\mathrm{T}/\mathrm{T}_0 \qquad \mathrm{T} < \mathrm{T}_0 \tag{1}$$

where

$$\mathrm{T}_0 = \frac{2h^2}{mk}\left(\frac{3n}{\pi}\right)^{\frac{2}{3}} \qquad \sim 50{,}000^\circ\ \mathrm{K}, \tag{2}$$

where n is the number of electrons per unit volume.

A term of this very small order of magnitude has actually been found by Keesom and Kok† for silver below 4° K (one electron per atom).

The breadth of the *d* band, however, is much less than that of the *s* band, because of the small overlap between the atomic *d* wave functions,‡ and therefore the density of states in the *d* band is very great. Thus, although the number of positive holes, 0·6, in the *d* band is independent of temperature below the melting-point, their distribution among the states is not, as may be seen from the disappearance of the spontaneous magnetization at about 360° C. Owing, moreover, to this large density of states, the heat capacity will be much larger than that given by (1) and (2).

Denoting by N (E) the density of states per gm/atom for a " positive hole " of energy E, and assuming that the holes obey the Fermi-Dirac statistics, we may write for their internal energy

$$\mathrm{U} = 2\int_0^\infty \frac{\mathrm{N}\,(\mathrm{E})\,\mathrm{E}\,d\mathrm{E}}{e^{(\mathrm{E}-\zeta)/k\mathrm{T}}+1},$$

where ζ is a function of T given by

$$2\int_0^\infty \frac{\mathrm{N}\,(\mathrm{E})\,d\mathrm{E}}{e^{(\mathrm{E}-\zeta)/k\mathrm{T}}+1} = \text{number of positive holes per gram atom.}$$

If we assume that N (E) is of the form $c\sqrt{\mathrm{E}}$, we may show that the heat capacity per electron is a function of a single parameter T_0; $k\mathrm{T}_0 = \zeta_0$

---

* Wigner and Seitz, ' Phys. Rev.,' vol. 46, p. 509 (1933); Fuchs, ' Proc. Roy. Soc.,' A, vol. 151, p. 585 (1935).

† ' Physica,' vol. 1, p. 770 (1934).

‡ *Cf.* Slater, ' Phys. Rev.,' vol. 36, p. 57 (1931).

is the maximum kinetic energy which any " hole " has at the absolute zero of temperature. T_0, however, will not be given by (2), but will have a much smaller value, corresponding to the large density of states in the *d* band. For $T \ll T_0$, formula (1) holds; for $T \gg T_0$, the heat per electron tends to the classical value $\frac{3}{2} k$; for intermediate values it can be obtained by numerical integration and is shown in fig. 2.

Comparing formula (1) with Keesom's results for nickel, we obtain $T_0 = 3100°$ K; this, however, is for ferromagnetic nickel, where all the spins of the " positive holes " point in the same direction, and therefore each electronic state is only singly occupied. Above the Curie point *two* electrons may be in any one state; since for $T = 0$

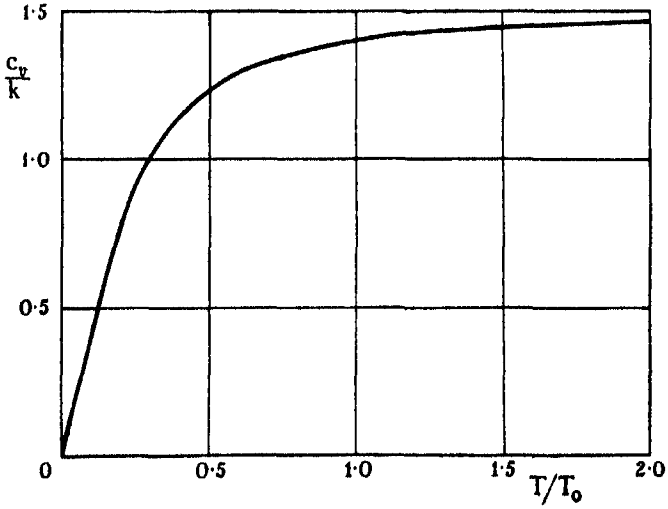


FIG. 2—Heat capacity c_v per particle of a Fermi-Dirac gas. k denotes Boltzmann's constant, T_0 the critical temperature.

$$\int_0^{kT_0} N(E)\, dE = \tfrac{2}{3} c\, (kT_0)^{\frac{3}{2}} = \text{number of occupied states with given spin direction,}$$

we see that the correct value of T_0 for *paramagnetic* nickel is

$$3100/2^{\frac{2}{3}} = 1950° \text{ K}.$$

At 600° K, T/T_0 has thus the value 0·32, and, as we see from fig. 2, the heat capacity per electron is about $1 \cdot 0\, k$. At this temperature, therefore, the contribution to the atomic heat of the metal will be

$$0 \cdot 6 \times 1 \cdot 0 \times R = 1 \cdot 2 \text{ cal/gm atom,}$$

and at 1300° K this will rise to 1·6 cal/gm atom; these correspond to the observed values within experimental error.

It appears therefore that the term observed by Keesom and Clark and the "unknown term" in the specific heat above the Curie temperature have one and the same origin.

Grew (*loc. cit.*) has measured the specific heat of a series of Cu-Ni alloys above the Curie temperatures, and finds that the excess specific heat diminishes with increasing concentration of copper, apparently tending to zero at the composition 60% Cu, where also the magnetism disappears. This is to be expected; according to the theory given by the author (*loc. cit.*), in an alloy containing copper and nickel in the ratio $x : 1 - x$, the number of positive holes is $0 \cdot 6 - x$ ($x < 0 \cdot 6$), their contribution to the atomic heat must therefore be less than

$$\tfrac{3}{2}\,\mathrm{R}\,(0 \cdot 6 - x) = 3\,(0 \cdot 6 - x) \text{ cal/gm atom.}$$

In this connexion it would be of great interest to know the atomic heat of these alloys at low temperatures, and to see whether there also the electronic specific heat tends to zero at the same composition.

Dr. KRONIG asked Professor Mott whether the excess specific heat, found in most metals, was much smaller than that obtained for nickel.

He also asked Professor McLennan whether any more work had been done recently in Toronto on high frequency currents in supraconductors.

Prof. MOTT replied that the excess specific heat for silver was 0·2 cal/gm atom, and that for most metals it appeared to be less than half that for nickel.

Dr. M. BLACKMAN—I should like to draw attention to the results on the specific heat of KCl given by Professor Keesom. As the temperature is decreased in the liquid air region the θ_D value first rises, then bends round, the top portion being more or less flat. This confirms the theoretical prediction made in some recent work on the theory of specific heat* that the rise in θ_D value would be followed by a region of constant θ_D value. This would seem to be the first case in which the true "continuum" region is reached.

---

* Blackman, 'Proc. Roy. Soc.,' A, vol. 149, p. 117 (1935).

Studies on Explosive Antimony. II—Its Structure, Electrical Conductivity, and Rate of Crystallization

By C. C. COFFIN, Ph.D., Dalhousie University, Halifax, Nova Scotia

(*Communicated by A. S. Eve, F.R.S.—Received January 29—Revised June* 7, 1935)

[PLATES 1 AND 2]

1—INTRODUCTION

The first paper* of this series dealt with the microscopically visible changes in structure that occur when the presumably amorphous electrolytic deposit of antimony "explodes." It was shown that a rapid (20–40 cm/sec) autogenous crystallization spreads spherically throughout the metal, leaving it with an onion-like structure visible on a polished and etched surface as a sequence of closely spaced (2000–3000 per/cm) concentric lines. The present paper is concerned with the nature of the amorphous deposit as revealed by a microscopic examination of etched surfaces, § 3, and by a study of its electrical resistance, § 4. The rate of crystallization at temperatures below that at which explosion occurs has also been determined, § 5.

It is found that the explosive deposits are characterized by a heterogeneous gel-like structure definitely oriented with respect to the cathode receiving surface. The electrical properties of the deposits are non-metallic in that the conductivity is very small and has a positive exponential temperature coefficient. Ohm's law is obeyed, but Faraday's law does not seem to be involved, although a small polarization is built up at the higher temperatures. At temperatures too low to initiate explosion the deposits crystallize at a rate which is independent of the extent to which they have already crystallized. The rate of crystallization increases exponentially with the temperature.

2—EXPERIMENTAL

The explosive antimony was deposited upon copper cathodes from 10% HCl solutions of commercially pure $SbCl_3$. Solutions containing less than 15% $SbCl_3$ were electrolysed at 0°. The more concentrated solutions (15–75% $SbCl_3$) were electrolysed at room temperature (19–21°). All depositions were thus well within the range of conditions favourable to the production of the explosive form.* The anodes of cast commerically pure antimony were carried on heavy platinum wires and were enclosed in linen bags to prevent possible contamination of the cathodes by anode slime.

---

* Coffin and Johnston, 'Proc. Roy. Soc.,' A, vol. 146, p. 564 (1934).

The specimens required for this work had to be removed from the copper without exploding and were usually desired in the form of long, narrow, and fairly thick bars. It was necessary also for the deposit to be as uniform as possible, *i.e.*, the temperature, current density, and particularly the $SbCl_3$ concentration at the cathode surface, had to remain constant throughout the deposition.

These requirements were fulfilled in a very satisfactory manner by the combination of cathodes and stirrer shown in fig. 1*a*. Four flat copper cathodes (7 × 1 × 0·14 cm) were fixed by set screws to a one-eighth inch horizontal brass plate carried by the stirring shaft. With the object of minimizing the centrifugal displacement of the solution, each cathode was offset so that the leading edge describes a larger circle than the trailing edge, fig. 1*b*. The whole assembly, with the exception of a 6 × 0·4 cm area on both sides of each cathode, was heavily painted with melted paraffin wax which, on the cathodes, was built up and roughly stream-lined, as shown in fig. 1*c*. With this apparatus working in a 600-cc beaker stirring speeds as high as 1000 r.p.m. may be obtained without spilling or splashing.

The electrolyses were run for 20–30 hours at current densities between 0·05 and 0·1 amps/cm^2. After washing with HCl solution, the cathodes were removed from the stirrer under water in order to prevent any local explosion from setting off the entire deposit. By carefully bending the copper strips under water, it was usually possible to separate the unexploded metal from the copper in the form of bars 1 or 2 cm in length. Antimony deposited from solutions containing between 10 and 20% $SbCl_3$ must be cooled in solid CO_2 before it can be safely stripped from the basis metal. Even at this temperature the greater part of the deposits from solutions of less than 10% $SbCl_3$ are lost.

The small rough bars of the explosive metal were then ground or filed to a rectangular cross-section. The deposits from solutions of more than 50% $SbCl_3$ may be safely filed under water. Those made from solutions of more than 20% may be ground by hand with fine emery and plenty of water on plate-glass at room temperature. Specimens from solutions containing as little as 10% $SbCl_3$ may be ground with emery powder and

---

† Cohen and Coffin, ' Z. phys. Chem.,' A, vol. 149, p. 417 (1930).

alcohol on a brass plate cooled in solid CO_2. The brass grinding-table shown in fig. 1*d* proved to be very satisfactory.

The surfaces to be etched were polished with fine rouge on wet chamois leather. Etching was done by making the polished specimen an anode in aqueous HCl or by immersing for 5–15 minutes in strong ammonium sulphide solution. Both methods give identical etch figures. The latter is the easier to control and was used in preparing the subjects of the accompanying photomicrographs.

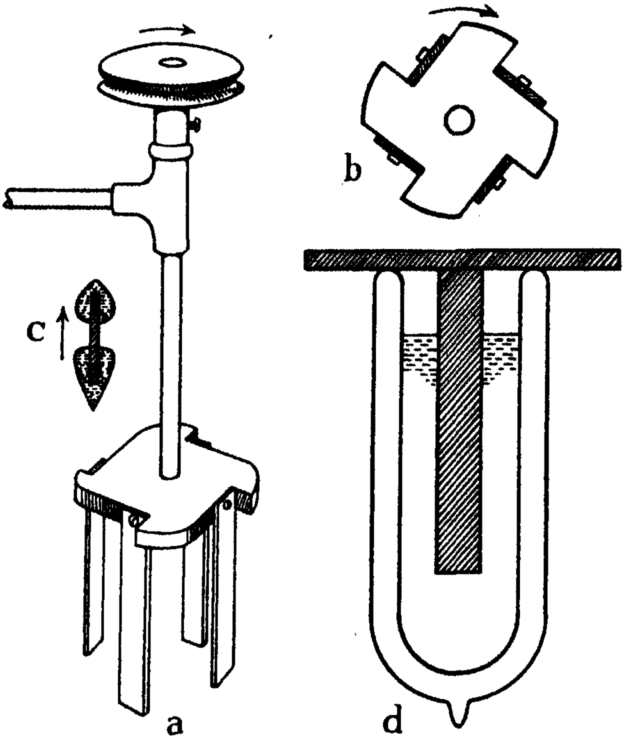


FIG. 1—*a*, *b*, and *c*, cathode stirrer; *d*, low-temperature grinding-table.

The small bars to be used for the resistance measurements were painted, except for a short distance from either end, with a quickly drying lacquer. Leads of No. 28 copper wire were wrapped several times about the bare ends of the specimen, which was then made a cathode in a silver-plating bath until a firm connexion of negligible resistance was established between the copper and the antimony. In cases where it was desired to determine the specific resistance of the specimen the dimensions of the painted (*i.e.*, not silver-plated) section were measured (before painting) with micrometer calipers. The specimens were then mounted as

" resistance thermomèters " in thin 5 mm glass tubes sealed off at one end and closed at the other with sealing-wax through which the leads passed. To determine the specific resistance of the deposits in the direction at right angles to their length (*i.e.*, parallel to the lines of the depositing current), small pieces were ground to rough cubes which were lightly silver plated. The silver was then ground off all but two opposite faces of each cube, which was measured up and fixed firmly in the amalgamated and insulated brass jaws of a small vice which could be slipped into a glass tube for temperature control.

Resistances were measured with a calibrated Wheatstone bridge.

3—THE MICROSTRUCTURE OF EXPLOSIVE DEPOSITS

The appearance of a polished and etched surface of explosive antimony is dependent upon whether the surface in question is parallel or perpendicular to the electrode face. If parallel to the cathode length, *i.e.*, normal to the stream lines of the depositing current, the etched surface is covered with a network of random markings somewhat like the crackle of pottery glaze, figs. 4, 5, and 8, Plate 1. If a surface perpendicular to the cathode face, *i.e.*, parallel to the current lines, is polished and etched a totally different type of marking is visible, figs. 6, 7, and 9, Plate 1, and figs. 10 and 11, Plate 2. In this case the etch figures are very definitely oriented in a direction parallel to the current lines. In the following the two types of etch figures will be called respectively " random " and " radial " markings.

It is at once apparent that the term " amorphous " should not be used in connexion with a material so definitely anisotropic as explosive antimony. The obvious contradiction loses a great deal of its significance, however, when it is remembered that the explosive deposits contain from 2 to 10% of $SbCl_3$* possibly in a crystalline condition.† Any anisotropism of the system as a whole may then be due to some preferred orientation of the salt crystals or salt metal complex embedded in the truly amorphous metal.

The external appearance of a deposit of explosive antimony is often such as to suggest a fibrous texture. The macroscopic stalactitic structure resembling certain geological growths (*e.g.*, limonite) made up of fibres radiating from a line is, for example, shown in the photographs published by Cohen and Ringer.‡ The central line is here the platinum wire upon which the antimony was deposited. Figs. 12 and 13, Plate 2, are microscopic views of two different types of cathode surface. The botryoidal structure usually due to the growth of fibres radiating from points is plainly visible. It should be pointed out that this globular structure is not characteristic of all deposits of explosive antimony. High-current densities and stirring speeds appear to inhibit its formation, and in the work being reported the great majority of the deposits were perfectly smooth and in appearance resembled polished steel.

---

* Cohen and Ringer, ' Z. phys. Chem.,' vol. 47, p. 1 (1904).

† Steinwehr and Schulze, ' Z. Physik,' vol. 63, p. 815 (1930).

‡ *Loc. cit.*

The extremely friable nature and the earthy fracture of explosive deposits with a high $SbCl_3$ content also indicates that there is an anisotropic texture characteristic of the material. A bar of the substance is readily crumbled in a way that suggests a bundle of short, not very cohesive, fibres. Such a fibre-like orientation parallel to the current lines has been observed also in several electrolytically deposited crystalline metals,* where it is evidently due to the growth of crystallites in the direction of highest conductivity.

That the radial figures parallel the current lines is particularly evident if a cylindrical deposit formed upon a copper wire be sectioned, polished, and etched. In such a deposit the markings are truly radial with the wire as the hub of the wheel. If the cathode is smooth and uniform in thickness the etch figures are everywhere parallel to a radius. If the deposit is markedly botryoidal in structure the figures change direction as the cathode grows and always end approximately normal to its surface. The appearance of such a section under the microscope is strongly suggestive of a iron filing " map " of a part of a magnetic field, *e.g.*, fig. 7, Plate 1. In this case the deposit grew from N. to S. No change could be observed in the direction of the etch figures from the centre to the edge of smooth cylindrical deposits formed upon copper wires rotated at a high speed about their own axis during the electrolysis.

The question whether the radial and random figures brought out by treatment with ammonium sulphide or with chloride ions are due to the solution of the $SbCl_3$ or to some such mechanism as a differential solution of the metal itself in different states of strain will be discussed in a later paragraph. It is obvious, however, from figs. 4–11, Plates 1 and 2, that the density of both the radial and random markings increases with the amount of salt in the metal. Figs. 4, 5, 6, and 7 are etched deposits from a 70% $SbCl_3$ solution and therefore contain about 10% $SbCl_3$. Figs. 8, 9, and 11 are deposits from a 15% solution and so contain about 5%

---

* *E.g.*, Glocker and Kaupp, ' Z. Physik,' vol. 24, p. 121 (1924); Bozorth, ' Phys. Rev.,' vol. 26, p. 390 (1925).

$SbCl_3$. All the surfaces were etched for the same length of time in the same ammonium sulphide solution. The markings are plainly larger and more numerous in the metal containing 10% $SbCl_3$. Attempts to "photometer" the negatives of such photographs by means of a Weston "photronic" cell have shown that there is a parallelism between salt content and amount of etching, but as yet have given no reproducible quantitative data with regard to the relative density of the etch figures on the different kinds of amorphous antimony.

Another type of visible heterogeneity is sometimes found in the explosive metal. Long, narrow, light-coloured bands extend throughout the deposit, as in fig. 15, Plate 2. (In this case the band runs about N.E. by E. and on the negative is very distinct.) These bands usually appear to be parallel to the electrode length and are therefore probably due to some irregularity in the deposition.

4—THE ELECTRICAL CONDUCTIVITY

The technique of preparing and mounting the specimens of explosive antimony for the resistance measurements has already been described. As the same resistance was obtained by the use of either alternating current (1000 ∿) and telephones or direct current and a galvanometer the latter method was finally adopted. That Ohm's law holds at all measurable voltages is shown by the fact that the resistance of a given specimen is independent of the bridge current.

The results of the specific resistance measurement at 0° (ρ_0) are given in Table I. The dimensions and actual resistance of each specimen are listed in columns 3, 4, and 5. Column 1 gives the $SbCl_3$ content of the solution from which the metal was deposited and column 2 the amount of $SbCl_3$ in the specimen as interpolated from the data of Cohen and Ringer (*loc. cit.*). For the first eight specimens the resistance was measured in the direction at right angles to the lines of the depositing current. The resistance of the last four specimens is that parallel to the current lines. As the dimensional error in these latter measurements is much larger than in those preceding, they are not taken into account in calculating the average ρ_0, column 7. The values obtained, however, are not greatly different from the others, so that it appears that the conductivity of explosive antimony is the same, or is at least of the same order of magnitude, in all directions.

It is to be noted that the conductivity increases with decreasing $SbCl_3$ content and that at 0° explosive antimony containing 10% $SbCl_3$ has about 200,000 times the resistance of ordinary antimony, $\rho_0 = 4 \times 10^{-5}$.

The effect of temperature change on the resistance of typical freshly made (*i.e.*, practically pure amorphous) specimens is shown in fig. 2. Provided that crystallization has not taken place to any appreciable extent a straight line results from plotting log R against 1/T, so that the resistance at any temperature can be represented by the equation

$$2\cdot3 \log R = \frac{\alpha}{T} - \beta,$$

where the numerical value of the constant β is dependent upon the dimensions of the specimen and the amount of $SbCl_3$ it contains. The value of α (slope of the lines in fig. 2) is about $2\cdot0 \times 10^3$ for freshly made deposits, irrespective of the amount of salt present, and remains constant down to liquid air temperatures where the resistance becomes too high to measure on a Wheatstone bridge, 10^7–10^8 ohms. If the specimen is not allowed to remain at room temperature for any length of time the resistance at lower temperatures is accurately reproducible. As the specimen ages at room temperature, however, and the degree of crystallization increases, the value of α diminishes and ceases to be constant. The curve of log R versus 1/T bends down and tends to become more and more nearly parallel to the 1/T axis as the temperature is lowered. In Table II are given some typical results of resistance measurements made on specimens aged at room temperature. It is to be noted that the resistance at −78° has decreased to a much greater extent than that at 0°.

TABLE I—THE SPECIFIC RESISTANCE OF EXPLOSIVE ANTIMONY AT 0°

% $SbCl_3$ in solution	% $SbCl_3$ in metal	Length cm	Cross-section cm^2	R_0 ohms	ρ_0	Mean ρ_0
75	10	1·55	0·0516	220·6	7·4	7·5
75	10	1·90	0·0312	457·4	7·5	
75	10	2·00	0·0875	173·8	7·6	
50	8	0·93	0·0623	68·7	4·6	4·7
50	8	0·68	0·0660	48·9	4·8	
15	5	1·35	0·0725	43·2	2·3	2·5
15	5	1·30	0·0973	34·0	2·5	
15	5	0·97	0·1038	25·7	2·7	
75	10	0·378	0·156	15·50	6·4	
75	10	0·263	0·101	13·61	5·4	
15	5	0·242	0·185	6·50	5·0	
15	5	0·180	0·145	4·70	3·8	

This decrease of α with increase of crystallization is presumably due to a tendency on the current lines to converge toward the crystalline

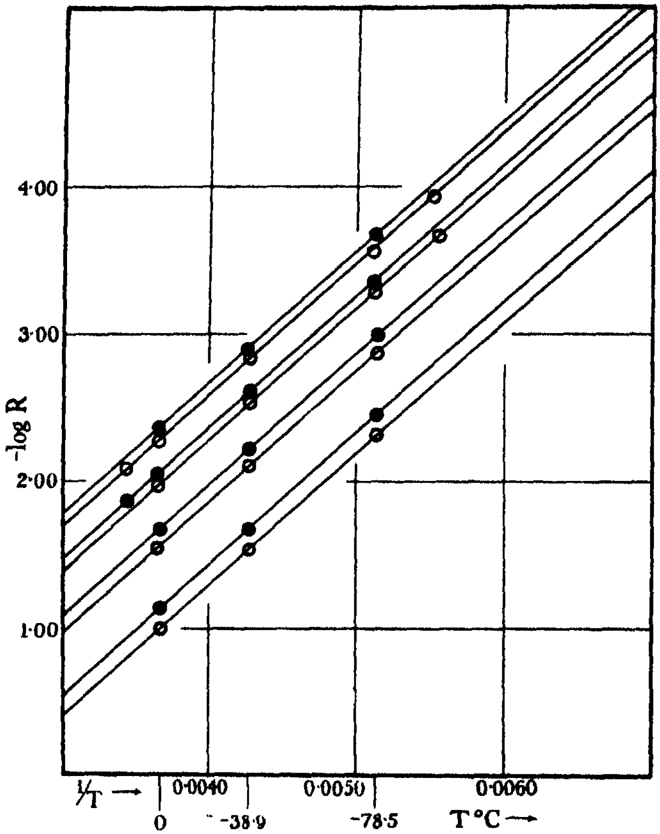


FIG. 2—Change of resistance with temperature.

TABLE II—EFFECT OF AGEING ON THE RESISTANCE OF EXPLOSIVE ANTIMONY

Age % $SbCl_3$	Several hours		4 weeks		8 weeks	
in metal	Ro°	R − 78°	Ro°	R − 78°	Ro°	R − 78°
10	220·6	4546	216·3	3895	199·0	2485
10	90·0	1838	89·3	1691	83·0	1192
10	33·5	7155	32·5	649·7	31·0	176·0
10	44·5	9401	43·3	759·9	39·0	555·0
5	—	—	43·2	571·7	35·5	395·0
5	—	—	34·0	472·7	32·0	431·0
5	—	—	25·7	324·9	20·7	139·0

regions as the conductivity of the surrounding medium diminishes, *i.e.*, as the temperature is lowered. In other words, the current density in the crystals increases and that in the amorphous matrix decreases as the temperature falls. Thus the tendency of the crystals to act " in parallel " and cause what might be called " internal short circuiting " increases as the temperature is lowered.

The assumption that there is a distortion of the current lines caused by a heterogeneity of the medium is supported also by the fact that the passage of a relatively small current results in the explosion of a partially crystallized specimen. Current densities of the order of 1 amp/cm^2 will explode a piece of the amorphous metal even if it is directly immersed in running tap water. If a specimen exploded in this way be polished and etched the concentric crystallization lines are plainly visible under the microscope and by their curvature may be followed back to a " crater " (occasionally more than one) somewhere in the interior of the metal. It thus appears probable that current densities sufficient to bring the temperature up to the explosion point (100°–120° C) obtain at one or more points within the specimen. It should be pointed out here that the explosive crystallization wave will still travel at room temperature in specimens already 60–80% crystallized. The speed of the wave decreases and the spacing of the explosion lines increases with the degree of crystallization.

As already stated, the value of β in the above equation depends upon the dimensions and $SbCl_3$ content of the specimen. For a cube 1 cm on the edge β has the values 5·31, 5·78, and 6·41 for explosive antimony containing respectively 10, 8, and 5% $SbCl_3$. As this relationship is roughly linear the specific resistance of any specimen of freshly made explosive antimony at any temperature may be approximately represented by the equation

$$\log \rho = \frac{870}{T} + 0{\cdot}095\,C - 3{\cdot}26,$$

where C is the percentage of $SbCl_3$ in the metal.

From the behaviour outlined in the preceding paragraphs it might appear that the conductivity of explosive antimony is electrolytic in character. The high specific resistance, the negative exponential temperature coefficient of resistance, and the fact that the above equation is the same as the one that fits the electrolytic conductivity of dielectric crystals* all point to this conclusion. Moreover, the fact that the values of α for the molecular type of crystal lattice, 3×10^3,* and for explosive antimony,

---

* Joffé, " The Physics of Crystals," McGraw-Hill Book Co., N.Y., 1928.

$2 \cdot 0 \times 10^8$, are of the same order of magnitude suggests that the mechanism of conduction is the same in each case.

There are other data, however, which indicate that the conductivity is purely metallic. For instance, no detectable transfer of matter occurred during several prolonged attempts at "electrolysis" in directions both parallel and perpendicular to the face of the deposit. Ohm's law is obeyed at all measurable voltages; the same value for the resistance is obtained by using either direct or alternating currents, while increase in salt content, *i.e.*, increase in the number of possible electrolytic carriers, actually lowers the specific conductivity. It is indeed difficult to imagine any ionization in a medium of such electrical and optical properties. It would seem, therefore, that in spite of a small polarization effect at the higher temperatures the conductivity is mainly if not wholly metallic in character, in keeping with the fact that the material is opaque and when polished exhibits the unmistakable metallic lustre.

The experiments of Steinwehr and Schultz* on the galvanic polarization of explosive antimony were repeated and extended. By means of a quickly operated double-pole double-throw switch a specimen connected across a measured voltage drop and kept at a constant temperature could be disconnected and thrown into a galvanometer circuit. The total resistance of this circuit was kept equal to the critical damping resistance (300 ohms) of the galvanometer—a 30-ohm instrument with a sensitivity of 1·5 microvolts per mm of scale.

At −78° no trace of a polarization E.M.F. could be observed after applying potential differences as high as 10 volts to the ends of a specimen for long periods of time. (Voltage drops greater than this cause the passage of currents large enough to heat the specimen.) At higher temperatures, however, measurable and reproducible galvanometer deflections were encountered, indicating that the material is polarized to a slight extent. At any one temperature this E.M.F. of polarization varies directly as the polarizing voltage, which has to be applied for 2–3 minutes before the galvanometer kick reaches a constant maximum. For any one voltage drop the polarization E.M.F. is a linear function of the temperature. Table III gives the results obtained with a typical specimen (No. 4 in Table I) of uniform cross-section. In column 2 are listed the polarization E.M.F.'s (in microvolts) caused by a potential drop of 1 volt per cm at each of four different temperatures. The linear relationship is evident. If left in the circuit the galvanometer returns to zero at a rate which is proportional to the magnitude of the deflection. The "half-life" of the polarization is given in column 3.

---

* 'Z. Physik,' vol. 63, p. 815 (1930).

The whole effect is very similar to that encountered in dielectrics and in all probability is due to an ionic displacement or orientation in the highly viscous medium. As is to be expected, the number of ions (polarization) and the ease of relaxation (ionic mobility) increase with increasing temperature.

TABLE III—POLARIZATION AND ITS RATE OF DECAY AT DIFFERENT TEMPERATURES

Temperature °C	Polarization volts × 10^6	$t\frac{1}{2}$ (sec)
22	202	24·7
0	147	28·9
−21	93	34·6
−78	0	—

5—THE RATE OF THE NON-EXPLOSIVE CRYSTALLIZATION

If a specimen of explosive antimony be kept at a temperature between 50° C and 100° C the electrical conductivity increases at a measurable rate. The higher the temperature the more rapid is the change. As has already been pointed out, an increase in conductivity at constant temperature is presumably due to an increase in the ratio crystalline phase/amorphous phase, so that resistance measurements may be used to follow the rate of crystallization. It is true that there exists at all temperatures a certain amount of what has been called " internal short-circuiting," so that a strict linear relationship between the concentration of the dispersed crystalline phase and the overall resistance of the specimen cannot be expected. However, as shown in Table II, the magnitude of this effect appears to increase exponentially with the difference between the specific conductivities of the amorphous and crystalline metal and to become less and less important as the temperature rises, *i.e.*, as the difference between the specific conductivities of the two phases diminishes. At the temperatures at which the resistance changes at a measurable rate there is thus, in all probability, no great error involved in assuming a linear relationship between the electrical conductivity of a specimen and the extent to which it has crystallized. The fact that the resistance changes at a rate which is proportional to its magnitude is indeed very good evidence that this assumption is justified.

Rates of resistance change were measured at 56·6° (acetone vapour), 80° (benzene vapour), and 100° (steam), and in the following are interpreted as rates of crystallization. The resistance always decreased according to the equation of a first-order reaction:

$$k = \frac{2 \cdot 303}{t} \log \frac{R_0 - r}{Rt - r}$$

in which r is the final resistance (leads plus crystallized specimen), R_0 is the resistance at the start, and R*t* is the resistance at time *t*. Velocity constants (k = slope × 0·434) were calculated from the slope of the straight line obtained by plotting $\log \frac{R_0 - r}{Rt - r}$ against *t* or by Guggenheim's method.* The results of these measurements of rate of crystallization are summarized in Table IV. The concentration of $SbCl_3$ in the solution from which the metal was deposited appears to have little or no effect upon the rate of crystallization. During the reaction $SbCl_3$ is liberated by the specimen which loses the fibrous characteristics of the explosive modification and becomes a definitely polycrystalline metal. A microscopic examination of polished and etched surfaces of the changed specimens has shown the crystallization to be perfectly random. No trace of the rhythmic orientation characteristic of the " exploded " metal is evident.

The slope (× 2·303 R) of the straight line obtained by plotting $-\log k$ against $1/T$ gives an activation energy of 27,300 calories per gm atom. The velocity constant of the crystallization at any temperature is given by the equation

$$\ln k = 31 \cdot 40 - \frac{27300}{RT} .$$

This 27,300 calories (1·2 electron-volts) which must be acquired per gm-atom before crystallization can occur is of the order of magnitude of an ionization potential. It may be that the acquisition of 27,300/N calories per atom actually results in a sort of ionization, *i.e.*, liberation of a " conductivity " electron, and so makes possible the growth of the ordinary metallic lattice.

Recognition of the large temperature coefficient of the rate of crystallization makes possible a clearer picture of the explosion phenomenon. If the heat evolved in the reaction is not conducted away rapidly enough to prevent a rise in temperature the effect will, of course, be cumulative and a purely thermal explosion will result. It is thus meaningless to speak of a definite explosion temperature characteristic of the substance.

---

* 'Phil. Mag.,' vol. 7, p. 538 (1926).

The same would seem to apply to the transition temperature observed by Kramer and Zahn* for the amorphous → crystalline change in thin metallic films.

The fact that explosive antimony crystallizes slowly at room temperature *throws doubt upon the reliability of previously determined physical* constants of the substance. An example from this laboratory will serve as an illustration. It was stated in Part I that the speed of the explosion wave was being determined by measuring the distance from each of two "craters" (one of which was made a known fraction of a second after the other) to the line of meeting of the two sets of waves (*see* figs. 11 and

TABLE IV—RATE OF CRYSTALLIZATION OF EXPLOSIVE ANTIMONY

% $SbCl_3$ *in solution*	R_0	$t_{\frac{1}{2}}$ (min)
	56·6° C	
60	28·0	363
40	28·9	377
60	44·5	350
40	15·5	325
60	40·5	380
40	9·85	380
15	9·74	360
15	9·37	350
15	19·6	373
75	75·4	353
		Mean = 363; $k = 3 \cdot 2 \times 10^{-5}$ sec^{-1}
	80·0° C	
60	25·2	23
40	28·8	24
60	15·1	21
40	4·88	19
75	40·0	20
75	26·0	20
75	7·02	23
		Mean = 21; $k = 5 \cdot 5 \times 10^{-4}$ sec^{-1}
	100·0° C	
40	10·8	3·2
15	1·57	3·0
15	19·6	3·2
75	26·0	2·4
75	23·2	2·3
		Mean = 2·8; $k = 4 \cdot 1 \times 10^{-3}$ sec^{-1}

---

* 'Naturwiss.,' vol. 43, p. 792 (1932); 'Z. Physik,' vol. 86, p. 413 (1933).

12, Plate 7, Part 1). The method was to drop two red-hot needles from different heights on to the polished specimen. The needles were released simultaneously from electromagnets connected in series. Distances of fall were measured with a cathetometer; " crater-to-crater " and " craters-to-line " distances were measured with a microscope and mechanical stage. Fairly consistent results (10 to 40 cm/sec), which showed that the speed of explosion depended largely upon the temperature, salt concentration, and thickness of the deposit, were obtained. The age of the specimens used ranged from several weeks to several months.

As soon as it was found that the deposits should be 50% crystalline after about a month at room temperature the experiments were repeated with specimens no more than several hours old. It was then found that under all conditions (of temperature, etc.) the explosion was so violent that the deposits were shattered and stripped from the copper or so distorted that microscopic measurements were impossible. Fig. 14, Plate 2, is a photomicrograph of a fragment of a thick specimen that was badly shattered by the explosion. The hot needle struck at the centre of the concentric cracks, and the manner in which the deposits are split up into spherical shells is evident. As is to be expected, the speed and violence of the explosion is greatly reduced by partial crystallization. The point stressed here is that other properties also must be affected by ageing.

The actual rate of crystallization at room temperature is, however, much slower than that calculated from the high-temperature data. The resistance at 0° of the specimens listed in Table II has diminished only 8–12% (instead of the calculated 60%) after 8 weeks at room temperature. It appears also that thin pieces crystallize more rapidly than thick ones, and that lacquered specimens crystallize more slowly than those from which the lacquer has been removed. Moreover, by grinding off the surface layers of a fairly thick partially crystallized specimen it is possible to raise its resistance temperature coefficient—in favourable cases α can be brought up almost to that ($2 \cdot 0 \times 10^3$) of the freshly made material. It would seem, therefore, that at the lower temperatures the crystallization tends to work inward from the surface and that its rate may be influenced to some extent by the rate of escape, *i.e.*, by the vapour pressure, of the $SbCl_3$. The fact that the crystallization at the higher temperatures is a precise first-order reaction indicates that it is here a uniformly random process—possibly because the escape of $SbCl_3$ no longer determines the rate. If, at these temperatures, the surface layers crystallized much more rapidly than the interior complications due to short-circuiting and dependence of velocity constant on the surface, volume ratio of the specimen would be expected. The fact that deposits of sublimed $SbCl_3$ are to be observed on the walls of stoppered bottles in which massive deposits have been stored for several years shows that the salt can and does leave the metal as it crystallizes at room temperature.

In this connexion it is a peculiar fact that a specimen partially crystallized by storage at room temperature will regain part of its resistance when kept at 0°. In fig. 3 the percentage change in the resistance at −78° is plotted against time for four typical specimens, the initial resistance of which ranged from 715 to 4546 ohms. During the periods covered by the continuous lines the specimens were kept at 20°; between the points connected by the broken lines they were kept in ice. An unmistakable uniformity of behaviour is evident. The resistance always diminishes at 20° and increases at 0°. The resistance of a fresh specimen does not increase when kept in ice (as is to be expected, a very slow decrease occurs), nor can the original resistance of an aged specimen be completely recovered by storage at 0°.

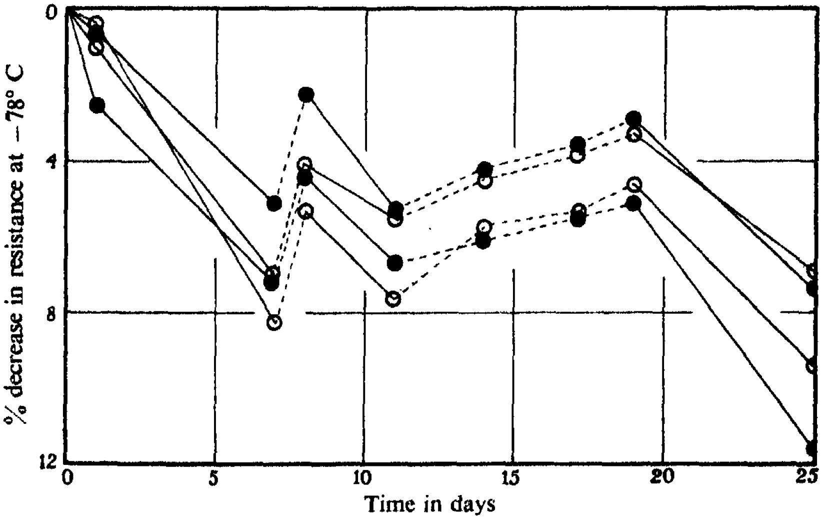


Fig. 3—The effect of storage temperature on resistance change.

Discussion

As has already been pointed out in Part I, there is good evidence for believing that the explosive deposits consist of amorphous antimony containing 2 to 10% $SbCl_3$. The part played by the salt in this peculiar form of matter is still obscure. The etch figures, Plates 1 and 2, may be due to increased reactivity along lines of internal stress in a homogeneous medium or they may be caused by a difference in the solubility of the two phases of a heterogeneous system. The peculiar growth and the oriented fibrous structure of the deposits suggest that the latter is the correct interpretation. While it seems probable that the amorphous metal constitutes one phase, the nature of the other, if indeed it exists, is not clear. Although the etch figures increase in density with increasing $SbCl_3$ content they cannot be due to the removal of $SbCl_3$ as such since they do not appear during prolonged immersion in such excellent $SbCl_3$ solvents as alcohol and ether and aqueous HCl. The hypothesis suggested for further investigation is that this second phase consists of an Sb-$SbCl_3$ complex resulting from the deposition of a complex ion. The important point at the moment is that explosive antimony is probably a heterogeneous " gel " and not a homogeneous solid solution of $SbCl_3$ in the amorphous metal as has previously been supposed.

It is questionable whether or not the explosive deposit should be referred to as a " metal." A consideration of the electrical conductivity —probably the most useful criterion of the metallic state—leads to no decision. Increase in the $SbCl_3$ content, however, increases the resistance, so that there is probably a small but real metallic conductivity characteristic of the amorphous phase; but amorphous antimony should not be likened to the super-cooled liquid. While both are " amorphous " the binding in the former must be predominantly homopolar and in the latter metallic.

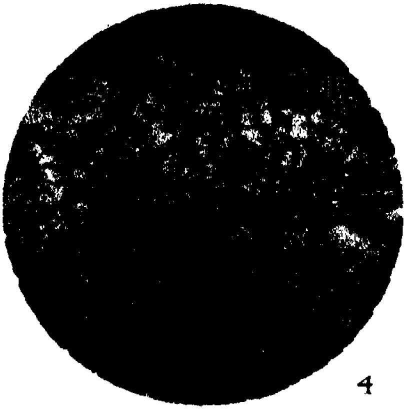
The writer wishes to acknowledge the assistance of Mr. Stuart Johnston and Mr. N. A. D. Parlee with part of the experimental work.

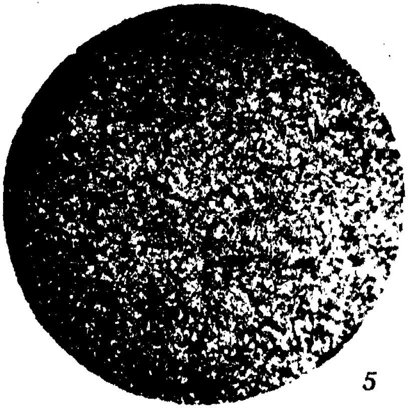
SUMMARY

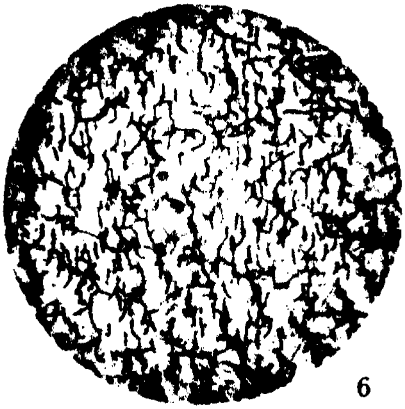
A microscopic study of polished and etched surfaces has shown that the explosive electrolytic deposit of antimony possesses a heterogeneous gel-like structure in which one phase is oriented parallel to the lines of the depositing current. This oriented phase, which is soluble in ammonium sulphide solution, is regarded as an $SbCl_3$-Sb complex resulting from the deposition of a complex ion. The other phase is probably amorphous antimony.

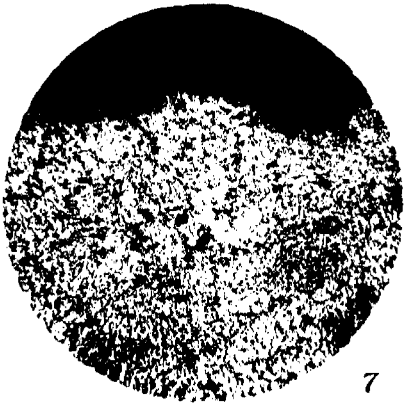
In the case of fresh deposits the specific resistance, which is about 10^6 times that of ordinary antimony and which has a negative exponential temperature coefficient, is given by the equation

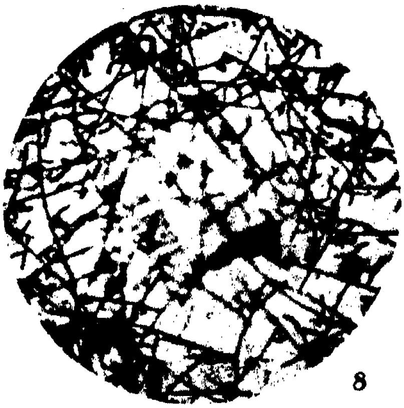
$$\log \rho = \frac{8 \cdot 7 \times 10^2}{T} + 0 \cdot 095\,C - 3 \cdot 26,$$

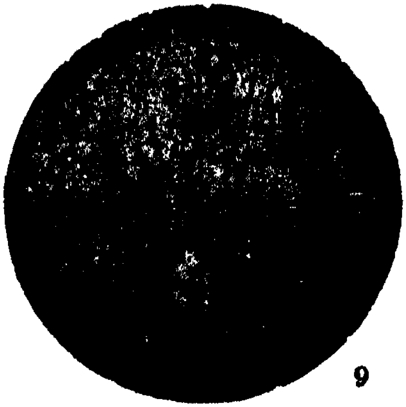

4

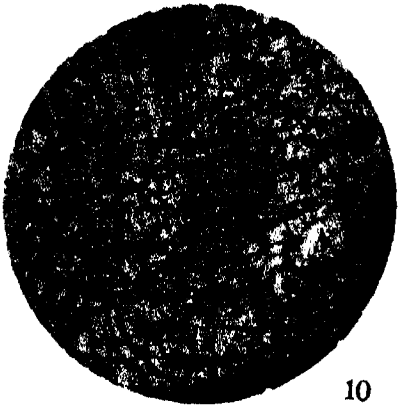

5

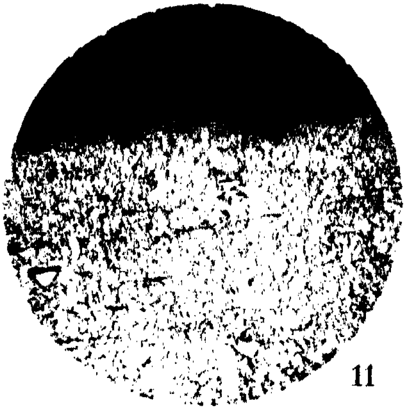

6

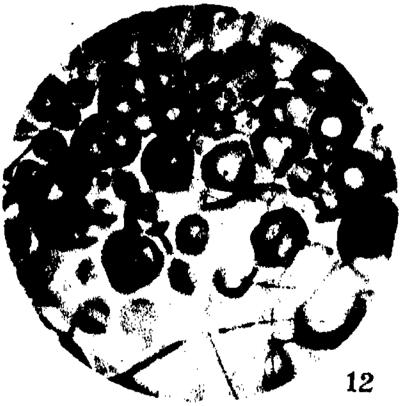

7

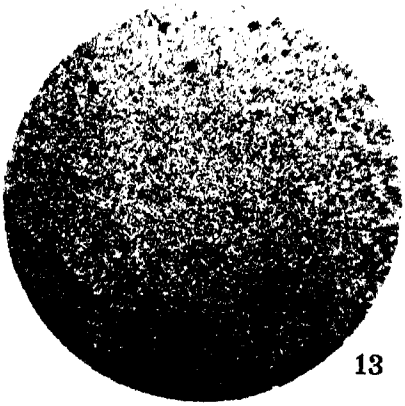

8

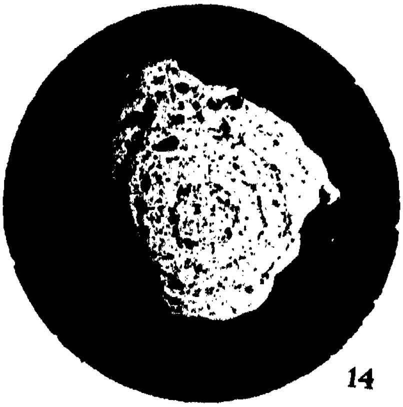

9

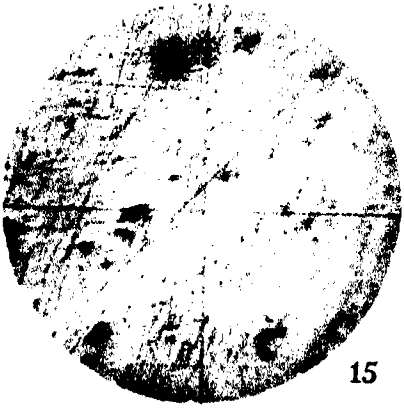

10


11


12


13


14


15

where C is the percentage of $SbCl_3$ in the metal. (This equation does not apply to partially crystallized deposits.) Ohm's law is obeyed and Faraday's law is not involved, so that, in spite of the resistance temperature coefficient and a small polarization which appears at the higher temperatures, the conductivity is probably metallic.

Between 55° C and 100° C amorphous antimony crystallizes at a measurable rate which is independent of the amount of salt in the metal and which can be represented by the equation of a first-order reaction. The Arrhenius equation applies and gives an energy of activation of 27,300 cal/gm atom. The rate of crystallization is given by the equation

$$\ln k = 31 \cdot 40 - \frac{27,300}{RT} .$$

EXPLANATION OF PLATES

PLATE 1

FIG. 4—Random etch figures on polished face of deposit from 70% $SbCl_3$ solution. × 180.

FIG. 5—Same as fig. 4. × 45.

FIG. 6—Radial etch figures on polished end of deposit from 70% solution. × 180.

FIG. 7—Same as fig. 6 except that deposit is from a 60% solution. × 45.

FIG. 8—Same as fig. 4 on deposit from a 15% solution. × 180.

FIG. 9—Same as fig. 6 on deposit from a 15% solution. × 180.

PLATE 2

FIG. 10—Same as fig. 6 on deposit from a 60% solution. × 180.

FIG. 11—Same as fig. 6 on deposit from 15% solution. × 45.

FIG. 12—Botryoidal structure on surface of deposit from a 70% solution. × 180.

FIG. 13—Same as fig. 12 on deposit from a 15% solution. × 45.

FIG. 14—Fragment of exploded deposit showing concentric cracks about crater. × 10.

FIG. 15—White streak in explosive metal. × 180.

The Optical System of the Disappearing Filament Pyrometer

By F. A. CUNNOLD, B.A., A.Inst.P.

(*Communicated by A. C. Egerton, F.R.S.—Received January* 31—*Revised July* 9, 1935)

INTRODUCTION

The principles of the disappearing filament pyrometer are too well known to need any description, but, in order to avoid confusion, it is advisable to define the terms which are used throughout this paper.

The "first objective" forms an image of the "source" in the "filament plane," an image of the filament and source is then formed by the " second objective" in the " second image plane." This image is examined by an " eyepiece system." The combination of the second objective and the eyepiece system constitutes the " microscope " examining the filament and the image of the source.

The cone of rays falling on the first objective from the source is the " entrance cone," while that forming the image of the source in the filament plane is limited by the " entrance aperture." The cone by which the phenomena in the filament plane are examined is limited by the " exit aperture," and the image in the second image plane is formed by the " exit cone."

The ratio " exit aperture/entrance aperture " is the " aperture ratio " of the instrument and the ratio " brightness of filament/brightness of image of source " is the " brightness ratio."

Worthing and Forsythe* measured the brightness ratio for various filaments and entrance apertures with a fixed exit aperture and concluded that it was not, in general, unity but could vary between about 0·9 and 1·7. This large variation in the value of the ratio was attributed to diffraction, by the filament, of light from the source.

Fairchild and Hoover† conducted a very comprehensive investigation of the optical effects occurring at the edge of the filament and showed that, with a round filament and reasonable values of the exit and entrance apertures, there might be a dark line just outside the edge of the filament and a bright line just inside the edge. By using flat tungsten filaments they were able to eliminate the bright line inside the edge of the filament, so they concluded that this was due to reflection, at the surface of the filament, of the convergent light from the source. The dark line outside the edge of the filament was attributed to diffraction and could be rendered invisible by the use of a smaller exit aperture than entrance.

---

* ' Phys. Rev.,' vol. 4, p. 163 (1914).

† ' J. Opt. Soc. Amer.,' vol. 7, p. 543 (1912).

As a result of their experiments they decided that it was necessary to formulate a new law of diffraction, viz., " of light incident on an obstacle in an image plane, only that component is diffracted which represents the diffracted rays radiating from the edge of the aperture." Thus, with the flat filament, all that was necessary to remove the visible effects at the edges of the filament and so get good disappearance was to reduce the exit aperture.

When using a round filament, however, it was first necessary to use such relative apertures that the effects of reflection and diffraction were equal and opposite and, at the same time, such an exit aperture that the two effects were not resolved. (The word " resolved " is presumably used here in its widest sense as one cannot strictly speak of resolution and resolving power when considering a dark line juxtaposed to a bright one on a bright background.)

These considerations considerably limited the possible apertures but, if the filament were of diameter 0·04 to 0·06 mm, the entrance aperture 0·13 radian, and the exit aperture 0·04 radian, perfect disappearance was possible. With the flat filament there was no visible effect of diffraction if these same apertures were used.

The only way these two statements can be reconciled is to conclude that under these conditions there is no visible effect of reflection when using the round filament. There were several such instances of apparent self-contradiction in the paper, and the fact that observations were made not in agreement with those of Worthing and Forsythe made it seem that a further study of the effects occurring in the optical pyrometer might be well repaid. The results suggest that the aperture limits might quite well be widened.

REFLECTION AT THE ROUND FILAMENT

Fairchild and Hoover, referring to diffraction, emphasize that "illumination of the shadow means a *loss of light from without the shadow*." They do not, however, seem to have applied this principle to reflection at the round filament for it indicates that the filament can only gain light by reflection if the image of the source loses light. Thus the bright line of reflection must be accompanied by a dark line; they make no mention of such a line. The following simple calculation shows that the dark line is at least as important as the bright.

Let us assume that the apertures of the pyrometer are squares such that the entrance and exit apertures are 2α and 2β parallel to the sides. Further, that the filament is straight and has its axis parallel to one side of each aperture and perpendicular to the optic axis—as in the pyrometer in use in the Thermodynamics Department, Clarendon Laboratory.*

Let RPS, fig. 1, represent a section of the filament such that the image of the source is formed in a plane through AO perpendicular to the plane of the paper. The optic axis is parallel to ROS and passes through the point P.

At a point A such that the tangent AB makes an angle α with the horizontal there will be no loss of light at the image of the source as the point image at A is formed by a complete cone of semi-angle α. But β may be less than α, so that there may be no visible loss of light until we are nearer P, *i.e.*, at C where the tangent CD makes an angle β with the horizontal.

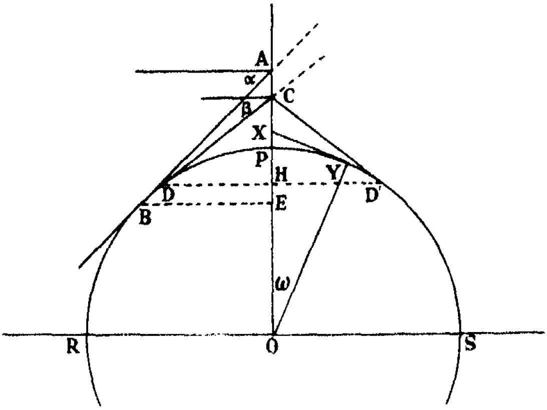


FIG. 1.

As we approach the point P from C the image will become darker and darker until at P there is zero illumination.

Take any point X on PC ($PX = x$) such that the tangent XY touches the filament at Y where the angle POY $= \omega$. The image at X is formed by an unobstructed cone of semi-angle ω, so that

$$\frac{\text{loss of light at X}}{\text{full brightness}} = \frac{\beta - \omega}{\beta}.$$

Consider now an element of area at X of length dx and of unit width perpendicular to the plane of the paper. The amount of light lost by this element, expressed as a fraction of the amount which would be present per unit area if there had been no obstruction, is

---

* Milford, Bracey, Cunnold, and Egerton, 'J. Sci. Instr.,' vol. 12, p. 80 (1935).

$$\frac{\beta - \omega}{\beta} \,.\, dx.$$

But we have $x = r(\sec\omega - 1)$, where $r = \mathrm{OP}$ = radius of the filament. Therefore loss of light by the element of area at X is

$$\frac{\beta - \omega}{\beta} \,.\, d\,[r(\sec\omega - 1)]$$

$$= r\,\frac{\beta - \omega}{\beta} \,.\, d(\sec\omega).$$

Then the total light lost by obstruction of the image between P and C, expressed in the same way, is

$$\int_0^\beta \frac{r}{\beta}(\beta - \omega) \,.\, d(\sec\omega)$$

$$= \frac{r}{\beta}\left\{[(\beta - \omega)\sec\omega]_0^\beta + \int_0^\beta \sec\omega \,.\, d\omega\right\}$$

$$\hat{=} \frac{r}{\beta}\left[-\beta + \int_0^\beta \left(1 + \frac{\omega^2}{2}\right) d\omega\right] = \frac{r\beta^2}{6}.$$

So that if the intensity of the unobstructed parts of the image is I the amount of light lost over the portion PC is $(r\beta^2/6)$ I.

If we consider rays not in the plane of the paper it is clear that the actual amount lost will be less than $(r\beta^2/6)$ I, but, as it will be shown later that even this value represents an amount of light in general too small to be appreciable, we can neglect the effect of rays not in the plane of the paper.

Turning attention to the gain of light by the filament, it is clear that no light can reach the exit aperture from a point below D on the side of the filament remote from the aperture. Light may, however, reach the aperture from any point on the side of the filament facing it.

A similar calculation to the preceding one shows that the amount of light which may possibly enter the exit aperture after reflection at the surface of the filament is $\frac{1}{2}r(\frac{1}{6}\alpha^2 + \frac{1}{6}\beta^2)$ I.

The amount which does actually enter the exit aperture is less than this first, because, after reflection, some of the rays are too divergent to enter, and secondly, on account of incomplete reflection at the filament. Thus if k is a positive proper fraction and ρ the coefficient of reflection the amount of light which does enter the exit aperture is

$$k\rho\,\tfrac{1}{2}r\,[\tfrac{1}{6}\alpha^2 + \tfrac{1}{6}\beta^2]\,\mathrm{I}.$$

To calculate the changes in intensity of the filament and the image of the source caused by this reflection it is first necessary to calculate the area over which the effect takes place.

Now

$$\mathrm{AP} \mathrel{\hat{=}} \mathrm{PE} \mathrel{\hat{=}} \tfrac{1}{2}r\alpha^2$$

and

$$\mathrm{CP} \mathrel{\hat{=}} \mathrm{PH} \mathrel{\hat{=}} \tfrac{1}{2}r\beta^2,$$

so that the whole effect is confined to a band, parallel to the filament, of width $\frac{1}{2}r\,(\alpha^2 + \beta^2)$ in the filament plane.

Usually α and β are each of the order 0·05 radian and r is about 0·05 mm, so that $\frac{1}{2}r\,(\alpha^2 + \beta^2) = 1{\cdot}25 \times 10^{-5}$ cm, *i.e.*, less than the wavelength of the light used to observe the effect. If we may speak of a reflection effect under these conditions it is quite certain that the separate effects (increase and decrease of intensity) will not be separately visible. However, there may be a resultant effect, giving a bright line, due to an amount of light

$$\tfrac{1}{12}\,r\,\{k\rho\,(\alpha^2 + \beta^2) - 2\beta^2\}.$$

A rough graphical determination of the constant k showed that it differed little from unity—only a few per cent.—for the normal values of α and β. Moreover, although ρ will be considerably below unity for the time being, we may assume that it is actually unity (thus exaggerating the effect).

Then the net gain of light by reflection is $\frac{1}{12}\,r\,(\alpha^2 - \beta^2)\,\mathrm{I}$.

This light is limited to a very narrow band parallel to the filament, so that to a first approximation we may assume that the image of it consists of a bright line of width equal to the diameter of the central disc of the Airy ring pattern for a point source, *i.e.*, a band of width $\lambda/2\beta$ approximately. The increase of intensity where this band falls is therefore

$$\frac{\frac{1}{12}\,r\,(\alpha^2 - \beta^2)\,\mathrm{I}}{\lambda/2\beta} = \Delta\mathrm{I}.$$

Fairchild and Hoover illustrate the reflection effect when $\alpha = 0{\cdot}05$ radian and $\beta = 0{\cdot}04$ radian, the diameter of the filament being 0·1 mm. Substituting these values in the equation just deduced we find $\Delta\mathrm{I}/\mathrm{I} = 5 \times 10^{-4}$, *i.e.*, the increase of intensity due to reflection is only 0·05% of the intensity of the field.

An effect of this order would not be visible and this figure has only been obtained by making assumptions such that the actual effect is considerably smaller. Thus if $\rho = 2\beta^2/(\alpha^2 + \beta^2) = 0 \cdot 78$, ΔI is zero, while if ρ is less than this, ΔI is negative, *i.e.*, there is a net loss of light due to the combined effects of obstruction and reflection.

In any case we may clearly conclude that the effect of reflection is negligible. Yet there is no question but that a bright line may be seen, under suitable conditions, just inside the edge of the filament. This line can be definitely attributed to the effect of deviations from Lambert's cosine law of emission on the part of the tungsten filament.

Worthing* showed that incandescent tungsten did not obey Lambert's law and that the emissivity was a maximum at an angle of emission of about 75°. From his results he calculated the brightness of a tungsten filament at various distances from the centre, assuming that it was viewed in parallel light. Near the edge the brightness was as much as 15% higher than at the centre, so that, under suitable conditions, a bright line would be seen just inside each edge.

Although Fairchild and Hoover attributed the bright line as seen in the optical pyrometer to reflection they also say " . . . This deviation is effective near the edges of the filament and is accounted for by the selection of such apertures that the dark band of diffraction, the bright band of reflection and the variable brightness near the edge at small angles of emission are balanced and the exit aperture is selected to just prevent resolution of the separate effects." This sentence is also worthy of note in that if the variable brightness near the edge of the filament and the bright band of reflection are mingled with, and balanced by, the dark band of diffraction, there seems no reason why there should be the true equality of brightness of filament and image which these authors assert.

It might be thought that the fact that the bright lines, as seen in the pyrometer, are more visible where the filament crosses the image of the source was a sufficient proof that they were caused by reflection. This effect is, however, due to the conditions of contrast for, where the filament is seen against the very dark background, in a very small distance the intensity falls from the maximum to zero, thereby masking the increase of brightness just before the final decrease.

In view of these rather confusing effects it seemed advisable to demonstrate experimentally as well as theoretically that the only difference between a round and a flat filament lies in the deviations from Lambert's law.

---

* 'Astrophys. J.,' vol. 36, p. 345 (1912); 'J. Opt. Soc. Amer.,' vol. 13, p. 635 (1926).

Visual observations were made using round filaments, flat filaments (made by rolling out fine platinum wire to a thickness of 5 μ), and also a razor edge, but no trace of a true reflection effect could be seen. About 40 photographs were also taken to show the combined effect of all the optical edge effects, but as these photographs were similar to those reproduced in Fairchild and Hoover's paper it is not essential to reproduce them here. Thus it is not in the actual appearance of the edge effects that the author differs from those observers but in the interpretation of them.

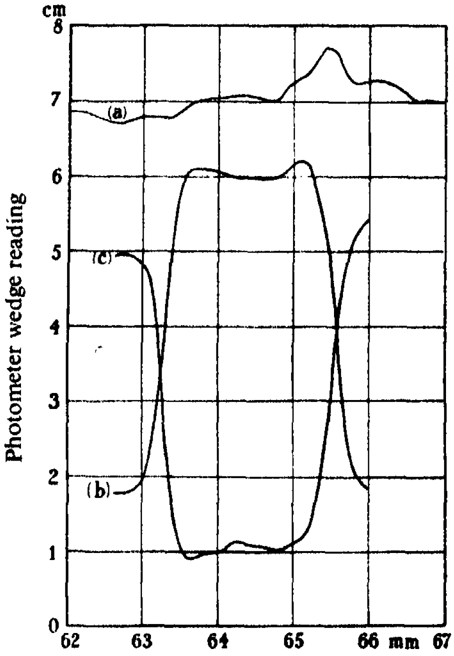


FIG. 2.

In order finally to settle the problem of the bright line a negative obtained by photographing the appearance of a bare platinum filament, 0·08 mm in diameter, with entrance and exit apertures 0·11 and 0·08 radian respectively was examined in a photometer. Fig. 2 shows the densities along lines crossing the image of the filament (*a*) where the filament crosses the image of the source and (*b*) just outside the image of the source. (It will be noticed that the effects of diffraction at the two edges of the filament, which are indicated by the steep portions of curve (*b*), are not similar; but this does not affect the results.) Curve (*c*) is the difference between curves (*a*) and (*b*) so that, if reflection is appreciable, there should be a rise of this curve just before the end of the lower horizontal portion. Curve (*b*) shows a well-marked effect due to the deviations from Lambert's law, but curve (*c*) shows no trace of the reflection effect.

It has therefore been demonstrated that reflection at the round filament plays a very minor part in the design of the disappearing filament pyrometer, which is contrary to the view previously held. Ribaud,* for instance, says: " On peut aisément percevoir cette lumière réfléchie sur les bords du filament en disposant au centre de l'objectif un diaphragme circulaire interceptant les rayons de la source qui arrivent à l'intérieur de l'angle β." It is certainly true that under these conditions there is a well-marked bright line at each edge of the filament; these, however, are due not to reflection but to diffraction.†

DIFFRACTION

In dealing with the problem of the diffraction effects seen in the optical pyrometer it seems advisable, first of all, to find out how these effects can arise. This is most simply done by separating the functions of the various optical components.

The first objective forms an image of the source in the filament plane, but, since the aperture is limited, the image of a point source is not a point but an Airy ring pattern. At a short distance from the edge of the image of the source there is complete overlapping of the ring patterns, so that we get uniform illumination; the filament simply obstructs a portion of this uniformly illuminated image.

The composite radiator, image, and filament, is next examined by another lens of limited aperture, *i.e.*, the second objective. Let us consider for a moment a single ring pattern which has originated from a single point of the source. We can treat each point of it as a point source giving rise to the usual ring pattern in the image plane. The separate patterns from these point sources overlap and since the light from each is coherent (having come originally from the same point source) an interference pattern will be produced. At a distance from any edge this interference pattern will be mingled with those from other ring patterns so that, once again, there will be uniform illumination. At the edge of the filament, however, there will be only incomplete overlapping owing to the obstruction of portions of the original ring patterns, so that there may be a resultant interference pattern.

---

* " Traité de Pyrometric Optique " (1931).

† Fairchild and Hoover, ' J. Opt. Soc. Amer.,' vol. 7 (1912), p. 566 and fig. 6G.

This final composite image is observed through an eyepiece, but, as there is no partial obstruction, no new interference effects can be introduced.

This analysis shows that there must always be some diffracted light or, rather, that an interference pattern must always be present. Thus the problem Fairchild and Hoover set themselves, *i.e.*, designing a pyrometer in which the effect is absent, cannot be solved except by making radical changes in design. It is, however, possible to reduce the magnitude of the effect until it is invisible.

Fairchild and Hoover considered that it was the difference of exit and entrance apertures that controlled the visibility of the effect and suggested theoretically that a difference of 0·02 or 0·03 radian was necessary for " . . . a negligible illumination of a minute exit aperture by diffracted rays radiating from the edge of the entrance aperture." Their experimental values are plotted in fig. 3, and it will be seen that, extrapolating to zero exit aperture, the difference is certainly not greater than 0·01 radian and might easily be zero.

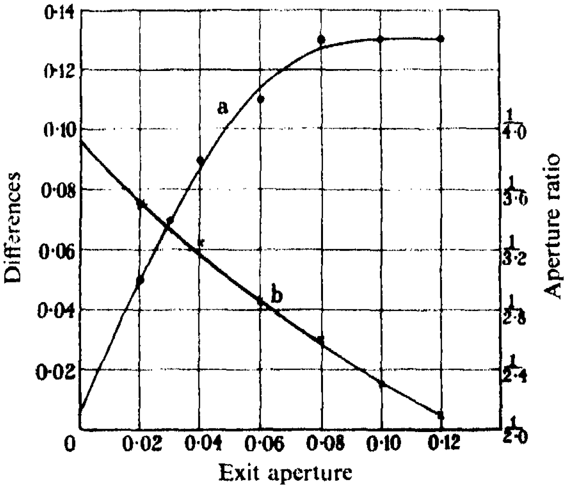


Fig. 3—● Difference curve; × ratio curve.

Since the interference pattern arises, as shown above, from the partial obstruction of the Airy ring patterns in the filament plane it is clear that there could be no interference effect if the image of the source formed in the filament plane consisted of point sources. It then seems reasonable to suppose that if the image of the source were composed of sources which were effectively points there would be little visible effect of diffraction. The Airy ring patterns may be effectively point sources if the resolving power of the pyrometer microscope is small, *i.e.*, if the exit aperture of the instrument is small, and, since the actual size of the ring patterns depends on the entrance aperture, it is to be expected that it is the relative sizes of these two apertures (the aperture ratio) that controls the visibility of the " diffraction " effect.

The values of the aperture ratios found by Fairchild and Hoover to be necessary for good disappearance are plotted in fig. 3, and it will be seen that the ratio is about ½ for large apertures but is considerably less for small apertures. These aperture ratios were checked by experiment and results in very good agreement obtained, so that there seems little doubt of their correctness.

As has been suggested, the interference pattern arises from the obstruction of portions of the Airy patterns in the filament plane so that, reducing the patterns to the central bright disc, the light giving rise to the interference pattern comes from a belt in the filament plane of width equal to the diameter of the bright disc and adjacent to the edge of the filament. The width of this belt is therefore inversely proportional to the entrance aperture, so that the brightness of the interference pattern will also be roughly inversely proportional to that aperture. Thus for the same visibility of the pattern the aperture ratio must decrease with decrease of entrance aperture, and this is actually found to be so in practice.

It is clear from the foregoing that the aperture ratio determines the character of the interference pattern since it fixes the extent to which overlapping of the various parts of the ring pattern occurs. The actual size, *i.e.*, the separation of the maxima and minima, of the pattern is determined by the exit cone; thus changing the exit aperture both changes the character and size of the interference pattern while changing the entrance aperture only alters the character of the pattern.

A complete mathematical treatment of the problem on these lines would probably enable the " diffraction " effects seen in the optical pyrometer to be predicted, but as such a treatment would be very complex it is not proposed to attempt it here. The following simple deduction, however, probably predicts the type of effect with reasonable accuracy and should serve to indicate the possibilities of the method.

Consider a portion of an Airy diffraction pattern; for simplicity take a half-pattern as far as the second dark ring, Divide this pattern into seven bands parallel to the diameter and select portions of these bands along the semi-diameter; fig. 4 makes this process clear. Let us regard each of these seven sub-divisions as a point source. The light from each of these point sources is coherent and we may assume that the phase over the central disc is constant* while that over the bright ring is always 180° different; moreover, the brightness of each source may be deduced from Airy's figures.† Each point source gives rise to a ring pattern in the conjugate plane of the examining lens (the second objective) and therefore, by summation, the light reaching a narrow band in this plane parallel to the semi-diameter on which the sources lie may be obtained. The distribution thus obtained shows little obvious change from that which would have been produced by a single point source at the centre of the original Airy ring pattern, *i.e.*, at O, except that the maximum has been displaced.

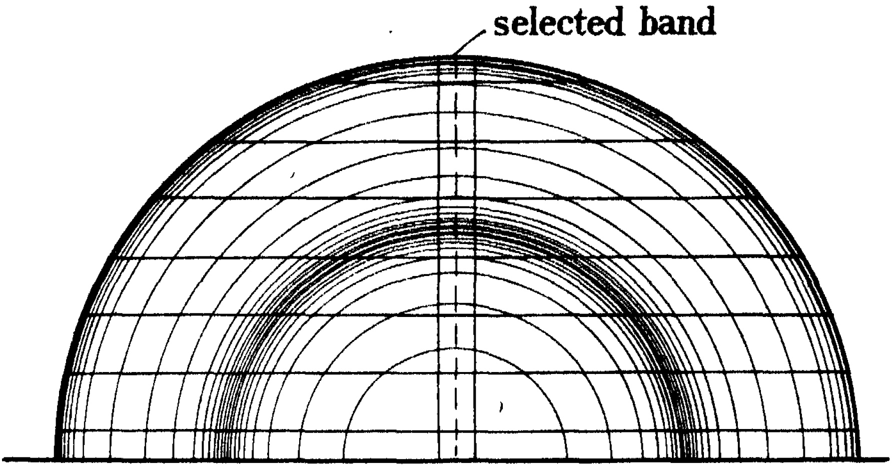


FIG. 4.

To compare the distributions more accurately the total intensity ($= \Sigma I$) due to each type of source (the seven " point " sources and the single one at O) was reduced to the same value and then the intensities due to the single point source were subtracted from those due to the seven. Where the resulting values were positive there should be an increase of illumination due to interference and where negative a decrease. Fig. 5 shows the distribution of light in the interference pattern, obtained in this way, for various values of the aperture ratio, *i.e.*, more or less overlap of the patterns from the seven point sources. The abscissae are given in radii of the first dark ring of the Airy pattern when the aperture ratio is unity and the exit cone is equal to the exit aperture. The ordinates are in per cent. of the intensity of the maximum of the single point source, while the image of the centre of the original pattern (O) falls at the point (O, O).

---

* Martin, ' Proc. Phys. Soc.,' vol. 43, p. 186 (1931).

† Airy, ' Trans. Camb. Phil. Soc.,' vol. 5, p. 283 (1834).

It will be seen that for the larger aperture ratios the dark band is deeper and narrower than the bright, so that it should be more visible, also that as the ratio increases the maxima and minima become closer and narrower. These features of the "diffraction" effect are exactly as found experimentally both by Fairchild and Hoover and the present author, so that the theory, in this case at least, predicts the type of effect with considerable accuracy.

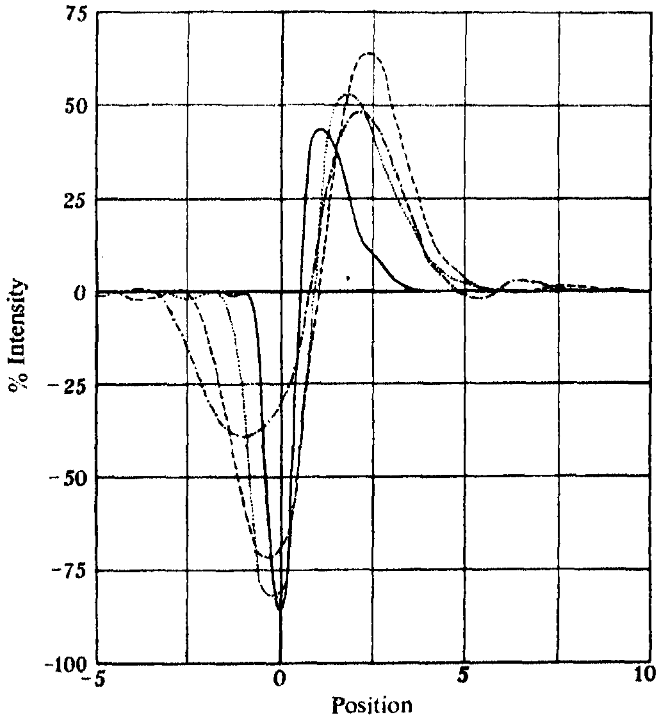


FIG. 5—Aperture ratio, ¼ —·—; ½ — — —; ¾ ....; 1 ——.

Unfortunately, however, such a simple deduction cannot possibly give any indication of the actual intensity of the diffraction effect, but it might be expected to give the separation of the maximum and minimum of the pattern with reasonable accuracy. To check this point two such separations were measured on photographs taken of the effect when the entrance aperture was 0·06 radian. When the exit aperture was also 0·06 radian (aperture ratio 1) the separation was equal to the radius of the central disc of the Airy pattern formed by the exit cone to an accuracy better than 10%. The calculation indicates that it should be very nearly equal to this radius, so that there is very good agreement, considering the many difficulties, both theoretical and practical. When the exit aperture was increased to 0·12 radian, thereby halving the size of the Airy pattern formed by the exit cone, the separation was also halved, as was to be expected.

The above simple deduction of the effects to be expected in the optical pyrometer is in such good agreement with the facts that there seems little doubt as to the justification of the general treatment of the problem. by considering the Airy ring patterns. Thus there is no necessity to formulate a new law of diffraction, as was done by Fairchild and Hoover, to explain the effects.

THE INFLUENCE OF THE EDGE EFFECTS ON ACCURACY

It is generally assumed that perfect disappearance of the filament in the disappearing filament pyrometer is essential for the following reasons.

1. Imperfect disappearance means that the brightness ratio is not necessarily unity.

2. If the disappearance is not perfect the edge effects are visible, so that there is a visible dividing line between the filament and the image of the source. Ordinary photometric practice indicates that the presence of this line should lead to a decrease of accuracy for one observer and inconsistent results for a number.

If we admit these reasons it is only possible to design a pyrometer with maximum accuracy if the limitations of the relative values of the apertures are in accordance with the data given by Fairchild and Hoover. Considering the first reason, however, it is clear that the brightness ratio need not be actually unity provided that it remains constant. This constancy can obviously be achieved by fixing the apertures, so that the edge effects are constant.* Turning now to the second reason, comparison with the presence of a dividing line in ordinary photometry is not reliable, as then the intensity of the line is usually very low, whereas in the pyrometer it may be only just visibly different in intensity from the filament and the source.

When actually making a pyrometer setting it is usual to adjust the filament rheostat until the filament goes first bright and then dark (or vice versa) against the background and then to set the rheostat midway between these two positions. The accuracy of the setting thus depends on the " bright-dark " range of the pyrometer; if this is small the instrument is accurate, if large inaccurate. The Fechner fraction, *i.e.*, the least perceptible difference of brightness of two sources expressed as a fraction of the mean brightness, is of the order 2%, so that the bright-dark range of a pyrometer with "perfect" disappearance is about 4%. Assuming this figure we might expect the accuracy of setting of such a pyrometer to be about 0·5% in brightness or about 0·5° in temperature at 1500° K.

---

* Worthing and Forsythe, 'Phys. Rev.,' vol. 4, p. 163 (1914).

Suppose, now, that a pyrometer were constructed so that there was a part of the filament 4% brighter than the rest; it would be possible to adjust the filament to one brightness only at which the bright part was just visibly bright against the background while the remainder was just visibly dark. That is, we have a pyrometer with a zero bright-dark range and therefore with maximum accuracy.

There seems no reason why, using suitable apertures, the bright lines due to the departures from Lambert's law should not, for the round filament, be made to introduce this rather brighter part of the filament. Accordingly observations were made through the pyrometer described elsewhere,* in which the fixed exit stop had been replaced by an iris diaphragm. This diaphragm therefore controlled the visibility of the bright lines due to the deviations from Lambert's law and, incidentally, the visibility of the effects of diffraction. It is clear that, for the test to be of value, all observations must be made at the same brightness as seen by the eye. To satisfy this condition a calibrated tungsten strip lamp was used as the source and its temperature varied with the exit stop. The characteristic of the pyrometer lamp had been determined approximately and it was found that the slope (dB/dC) of the brightness (B)-current (C) curve was approximately proportional to B/C, *i.e.*, $dB/B \propto dC/C$. The entrance aperture was fixed at 0·083 radian and the exit aperture could be varied between 0·02 radian and 0·07 radian.

Observations, using the special filament rheostat,† were made by three observers, of which two (M. M. and F. A. C.) had had considerable experience of pyrometry while the third (F. S.) had had none. With any single exit stop each observer made six settings of the filament current required to match the filament and the image of the source. From these, after correcting for any variations of the current flowing through the strip lamp source, the mean setting error for each observer was obtained and also the difference between the mean points of the different observers. The figures thus obtained, expressed as a percentage of the mean current, are given in Tables I and II, each of which contains the results for one filter only—observations were made using both a red and a green filter.

---

* Milford, Bracey, Cunnold, and Egerton, 'J. Sci. Instr.,' vol. 12, p. 40 (1935).

† Cunnold and Milford, 'J. Sci. Instr.,' vol. 11, p. 265 (1934).

These results show a distinct increase of accuracy with increase of exit aperture and consequent decrease of "perfection" of disappearance. It will be seen that the minimum value of dC/C is about 0·025% when the exit aperture is about 0·06 radian. This value corresponds to an error of about 0·2° at 1500° K, and is an improvement on the "photometric" accuracy of 0·5° at 1500° K.

TABLE I—RED FILTER

Exit aperture	Mean errors %					Mean differences %	
	F. S.	F. A. C.		M. M.	Mean	F. S. − F. A. C.	M. M. − F. A. C.
0·02	(0·09)	0·04				(+0·20)	
	0·04	0·06	0·03	(0·10)	0·04	−0·08	(+0·13)
0·03	0·04	0·05	0·04	0·03	0·04	−0·11	−0·02
0·05	0·03	0·05	0·02	0·02	0·03	−0·10	0·00
0·06	0·03	0·03	0·05	0·02	0·03	+0·03	−0·01
0·07	0·03	0·03	0·04	0·03	0·03	−0·07	+0·04

TABLE II—GREEN FILTER

Exit aperture	Mean errors %					Mean differences %	
	F. S.	F. A. C.		M. M.	Mean	F. S. − F. A. C.	M. M. − F. A. C.
0·02	0·03	0·02	0·05	0·06	0·04	−0·02	0·00
0·03	0·02	0·04	0·04	0·05	0·04	−0·08	+0·01
0·05	0·04	0·03	0·02	0·05	0·03	+0·07	−0·01
0·06	0·03	0·01	0·02	0·02	0·02	−0·02	+0·02
0·07	0·02	0·02	0·02	0·04	0·02	+0·02	0·00

We can therefore conclude that perfect disappearance of the filament in the optical pyrometer is not essential and, occasionally at least, its attainment involves a noticeable decrease of accuracy.

APPLICATION OF THE RESULTS TO THE DESIGN OF PYROMETERS

The conclusions arrived at by Fairchild and Hoover were as follows: "The precision, accuracy, and wide range of usefulness of the disappearing-filament optical pyrometer depend on perfect disappearance of the filament with high resolving power and magnification by the eyepiece. Older forms have not accomplished this. Non-disappearance of the filament is caused by diffraction and, if the filament is a wire, by reflection at its edges. Perfect disappearance of the flat filament with sharp edges is obtained by using an entrance aperture sufficiently large compared to the exit aperture. A round filament is made to disappear by decreasing the entrance aperture until diffraction and reflection nearly balance, in their effects at the filament." The following comment may be made on these conclusions: if the condition given in the first sentence is essential then the method of using a round filament, as conveyed in the last sentence, cannot be satisfactory.

As has been mentioned in the early part of the present paper, there is no simple relation between the resolving power of the microscope system (this is, presumably, referred to as the eyepiece in the above extract from the paper by Fairchild and Hoover) and the visibility of the diffraction and " reflection " effects taken together. The magnification of the microscope is the factor which is mainly effective in controlling this visibility.

The present work has shown that the so-called reflection effect is, in reality, due to the deviations from Lambert's law, so that if the resolving power of the microscope is equal to, or smaller than, the diameter of the filament the visibility of the bright lines at the edges of the filament depends very largely on the magnification of the microscope. If the exit aperture of the pyrometer is 0·02 radian the resolving power of the microscope system is about 0·04 mm, so that, since the filament of a pyrometer lamp is not usually smaller than this and the exit aperture is not usually less than 0·02 radian, it is the magnification of the microscope that usually controls the visibility of the edge effects. It has been shown that if the filament is 0·075 mm in diameter the exit aperture may be as high as 0·07 radian and the microscope magnification 10. It should therefore be possible to use this same magnification and an exit aperture as high as 0·10 radian if a 0·05 mm diameter filament is used. Also if the magnification is reduced to 7·5 it should be possible to use a filament of diameter 0·1 mm with the same exit aperture of 0·07 radian.

The conditions given by Fairchild and Hoover do not permit of the use or of an exit aperture greater than 0·04 radian with a 0·05 mm filament greater than 0·02 radian if a 0·1 mm filament is used. It will thus be seen that the present work considerably widens the limits imposed by Fairchild and Hoover, enables pyrometers transmitting much more light to be constructed, and also indicates that the round filament, instead of being inferior to the flat filament, may be superior. The importance of being able to design pyrometers transmitting more light is that (*a*) more truly monochromatic light filters can be used, and (*b*) the optical scale may be extended downwards, and it might then be possible to obtain a value of Wien's constant c_2 in terms of the optical scale.

In conclusion, the author would like to express his thanks to both Mr. M. Milford and Mr. F. Stanger for their assistance in making the observa-

tions on the accuracy of the pyrometer and also to Mr. A. C. Egerton, F.R.S., under whose direction the work has been carried out.

SUMMARY

It is shown, both theoretically and experimentally, that reflection at the surface of the round filament is negligible in the usual type of disappearing filament optical pyrometer. The bright lines which may be seen at the edges of the filament are attributed to the deviations from Lambert's cosine law.

A tentative theory of the diffraction effects at the edge of the filament is suggested and a simple deduction based on the same principles is shown to be in good agreement with the facts.

Measurements made with pyrometers in which the disappearance of the filament may or may not be perfect indicate that perfect disappearance is not essential for the attainment of maximum accuracy.

The influence of the edge effects on the design of optical pyrometers is discussed and it is shown that the aperture limits suggested by previous workers may be considerably widened so that pyrometers transmitting more light may be constructed and the optical temperature scale thereby extended downwards.

Numerical Relationships in Binary Metallic Systems

By D. STOCKDALE, King's College, Cambridge

(*Communicated by Sir William Pope, F.R.S.—Received March* 11, *Revised July* 29, 1935)

1—INTRODUCTION

The equilibrium diagrams of a family of binary alloys such as those of copper or silver with zinc, aluminium, or tin at first sight appear to be extraordinarily similar, but all efforts to predict quantitatively the silver-zinc diagram, for example, from a knowledge of the copper-zinc series and the properties of the silver, copper, and zinc atoms have failed. In this paper, certain relationships which are believed to exist in binary metallic systems are pointed out. There is also evidence, which is not now brought forward, to show that these relationships may also be found in binary systems of other types. If these simple relationships can be established, it will be at least a step towards the solution of the larger problem.

When a metal is added to copper, silver, or gold, it often forms a primary solid solution with the solvent element. On the addition of more of the solute, a two-phase range is reached, then a second or β-solid solution, which is often stable only at high temperatures, then another two-phase range, then a secondary or γ-solid solution, after which the phase relationships often become extremely complicated. The relevant portion of the copper-aluminium series is shown in fig. 1.

Hume-Rothery* has pointed out that the β-phases always contain that alloy where the ratio of atoms to valency electrons is 2: 3, *e.g.*, the alloys CuZn, Cu_3Al, Cu_5Sn. This Valency Rule was later extended to cover the γ-secondary solid solutions; there the ratio 13 atoms to 21 valency electrons is apparently of the greatest importance and the phases contain alloys of the formula, Cu_5Zn_8, Cu_9Al_4, $Cu_{31}Sn_8$. Owing largely to the work of Westgren and Phragmén,† and Bradley and Thewlis‡ and their collaborators, it has been established by the X-ray method that the β-alloys frequently belong to the body-centred cubic system and that the γ-alloys are also cubic, but with 52, or some simple multiple of 52, atoms in the unit cell. H. Jones§ has recently shown that alloys of the γ-type are to be expected when 52 atoms provide between 80 and 90 loosely-bound electrons. The Valency Rule requires 84. It is thus possible that certain simple numerical relationships between atoms and electrons or between solute and solvent atoms may determine not only the existence of a phase but also its structure. It is the object of this paper to show that at certain definite temperatures the appearance of a new phase is also determined by simple numerical relationships between solute and solvent atoms.

---

* 'J. Inst. Met.,' vol. 35, p. 295 (1926).

† 'Phil. Mag.,' vol. 50, p. 311 (1925), *inter alia*.

‡ 'Proc. Roy. Soc.,' A, vol. 102, p. 678 (1926), *inter alia*.

§ 'Proc. Roy. Soc.,' A, vol. 144, p. 225 (1934).

In a further contribution, in which the effect of solute atoms on the freezing- and-melting points and solid solubility limits of silver and copper alloys was exhaustively examined, Hume-Rothery, with Mabbott and Channel Evans,[*] showed that Raoult's Freezing-Point Law was not applicable to these systems. In its stead they put forward three rules:

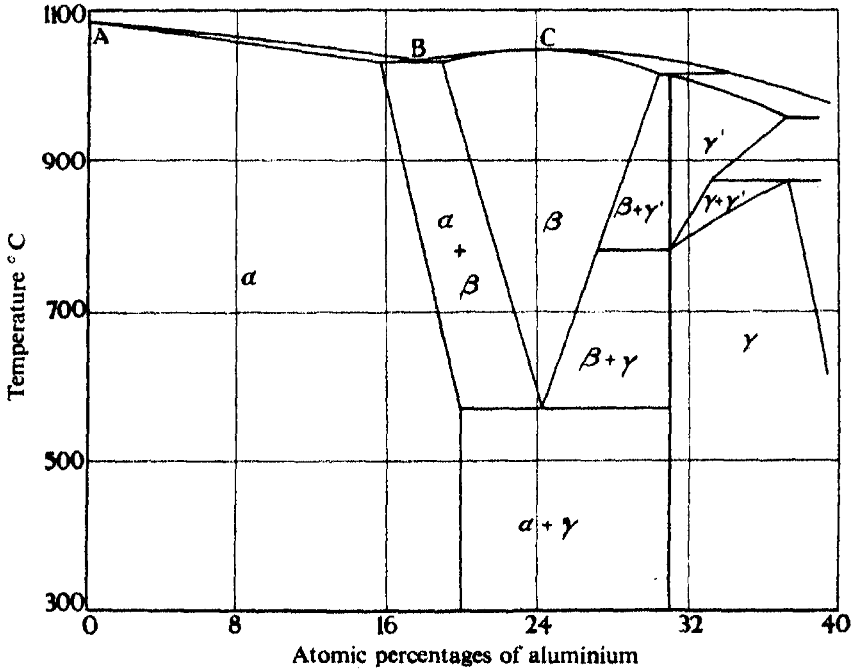


FIG. 1.

(i) The atomic compositions of alloys of a given freezing-point are inversely proportional either to the valencies of the solute elements or to a simple multiple or fraction of the normal valencies.

(ii) When proceeding along the same horizontal row of the periodic table, the atomic compositions of alloys of a given melting-point are approximately inversely proportional to the squares of the valencies.

---

[*] 'Phil. Trans.,' A, vol. 233, p. 1 (1934).

(iii) Where the atomic size-factors are favourable, the maximum α-solid solubility, to a first approximation corresponds to a constant electron concentration.

These rules have been adversely criticized,* but in my opinion the first is often true for dilute solutions. I prefer the rule in the alternative form that "The atomic depression of freezing-point is approximately proportional to the valency of the solute." I have considered the data given by the authors and find that out of 25 systems, in 19 at a concentration of 5 atomic %, a constant is obtained by dividing the depression of freezing-point (± 2° C) by an integer. If an experimental error of 3° C is permitted 3 other systems are brought in. The 3 remaining systems are those of lead with copper, and aluminium with both copper and silver. It is well known that the properties of aluminium in the metallic state are frequently anomalous, while the divergence in the copper-lead system is 4° C. I have taken 7 as the integral factor for antimony in copper; not 6, that chosen by the authors. In 10 of the best 22 systems, the integer is the valency; in the other 12 it is 3, 4, 5, 6, or 7. That the integer and the valency are identical in so many of the systems may be a coincidence.

The second rule is of less wide application than the first and the data in its support are less convincing. In both the copper and silver series the solidus lines when drawn to the appropriate scale fall rather into two groups than into one, but the impression is given that some simple relationship does exist between the solidus curves, though that relationship has not yet been exactly stated. These rules cannot be reconciled with the Van't Hoff equation for the depression of freezing-point in series which form solid solutions unless the existence of compounds between the solvent and solute in both the liquid and the solid states is postulated. The X-ray method has as yet failed to show the existence of such compounds in primary solid solutions. In this connexion the work of Jeffery† should be consulted. Reference to the third rule will be made later.

The importance which Hume-Rothery and his colleagues attach to the size-factor has not been sufficiently brought out in what has been written above, but it is evident that this factor, within certain limits, is thought by them to be subsidiary to valency. Formerly, I have taken the opposite view, and have been unable to advance beyond the elementary picture of a binary alloy crystal as something built up of balls of two different sizes. On this crude analogy it is the size-factor and above all the ratio of the numbers of atoms of different kinds which are important. A simple

---

* Anon, 'Metallurgist,' October, 1934.

† F. H. Jeffery, 'Trans. Faraday Soc.,' vol. 26, pp. 86, 587 (1930); vol. 27, p. 136 (1931).

numerical relationship between atoms of different kinds would necessarily lead to a simple atom-valency electron relationship.

My attention was first drawn to the importance of this numerical ratio by some work I was doing in 1924 on the copper-rich aluminium-copper-tin alloys.* It so happened that when considering the liquidus curves of this system, I plotted the sum of the atomic percentages of the solute elements, aluminium and tin, against temperature for each series of alloys containing a constant percentage of tin by weight. The resulting curves, within the limits of experimental error, which may be considerable owing to the high temperatures, some 1000° C, at which these alloys freeze, and to the flatness of the curves, were all parallel to the aluminium-copper liquidus A B C, fig. 1, and the line joining the pseudo-eutectic points was straight and parallel to the temperature axis. That is to say, the eutectic compositions over the range examined are determined by the ratio of solvent to solute atoms, irrespective of whether the solute is aluminium only or whether a considerable proportion of the aluminium is replaced by tin. The results are set out in Table I.

TABLE I—COMPOSITION OF ALUMINIUM-COPPER-TIN PSEUDO-EUTECTICS, IN ATOMIC PERCENTAGES

Copper	Aluminium	Tin	Aluminium and tin
82·2	17·8	Nil	17·8
82·0	17·5	0·48	18·0
81·9	17·1	0·97	18·1
82·2	16·3	1·46	17·8
81·5	16·0	2·48	18·5
81·7	14·7	3·56	18·3
81·5	13·9	4·62	18·5

The mean of the figures in the last column is 18·1 (4). As a ratio of 9 solvent atoms to 2 solute atoms requires an atomic percentage of 18·18, it was reasonable to consider the speculation that in all eutectic mixtures the atoms are present in a simple ratio.† This speculation recalls the original work of Guthrie,‡ who thought that cryohydrates were chemical compounds, and also that of Gorboff, who stated that if two components, A and B, form the compounds AB_x and AB_y, which melt without decomposition, the eutectic mixture of A and AB_x has the composition A + AB or A_2B_x and that of the two compounds has the composition $yAB_x + xAB_y$ or $A_{x+y}B_{2xy}$.*

---

* Stockdale, 'J. Inst. Met.,' vol. 35, p. 181 (1926).

† In any one piece of binary alloy there must necessarily be an integral number of atoms of the one sort with an integral number of the other. The word "simple" is possibly the only word available to express the idea that these numbers are divisible by a large common factor, with the result that their ratio can be stated in terms of comparatively small numbers.

‡ 'Phil. Mag.,' vol. 49, p. 1 (1875), *inter alia*.

A further speculation was made about the nature of primary solid solutions. There must be some kind of dynamic equilibrium between the solid and liquid in a freezing alloy and an interchange of atoms must take place between the two phases. After a certain arrangement of atoms in the solid phase has been reached, it may well be that the reaction of the solid on the liquid may change abruptly, so bringing about a change of phase in the liquid. Thus, suppose the crystal structure of the primary solid solution to be face-centred cubic, with solute atoms taking the place of solvent atoms on the lattice, there may be a certain relationship between the solid and the second phase until every eighth atom of solvent has been replaced, and a quite different relationship after the ratio has exceeded 1 : 7. This simple picture, which can no longer be upheld, led to the conclusion that at critical points on the equilibrium diagram the solute and solvent atoms would be in a simple ratio. Critical points are defined as those points which represent a uniform alloy in such a state that a small increase in solute will lead to the formation of a new phase, while a small increase in solute together with a small change in temperature will lead to the formation of a second new phase. The ratios to be expected with a cubic solvent would be simple ones, such as 1 : 15, 1 : 7, 1 : 3.

The literature was therefore examined and considerable support was obtained for these two speculations. However, as papers on alloys were continually being published which gave results neither in agreement with the earlier work nor with these theories, and as these questions seemed to be of importance, I embarked on a considerable programme of experimental work in order to obtain data. I chose six binary systems in which it was probable that serious experimental difficulties would not arise, and I measured the eutectic composition and the limits of primary solubility in these systems, and also investigated other points of importance. The results of the work on the composition of eutectics,† and on the equilibria in the cadmium-zinc, copper-silver, lead-tin, and copper-aluminium systems have been published‡ in the 'Journal of the Institute of Metals.' I have also examined the antimony-silver and cadmium-tin series, but the

* A. Gorboff, 'J. Russ. Phys. Chem. Soc.,' vol. 41, p. 1241 (1909).

† 'J. Inst. Met.,' vol. 43, p. 193 (1930).

‡ 'J. Inst. Met.,' vol. 44, p. 75 (1930); vol. 45, p. 127 (1931); vol. 49, p. 267 (1932); vol. 52, p. 111 (1933).

results obtained were largely confirmatory of earlier work and did not seem to be of sufficient general interest to warrant independent publication. In spite of the fact that this programme has taken seven years to complete, the data obtained which could be used in building up a theory of simple ratios were meagre, but fortunately the recent appearance of the paper by Hume-Rothery, Mabbott, and Channel Evans* has made a mass of information available. Not only does their own experimental work appear to be invaluable, but they also quote the results of other investigators. By referring almost exclusively to their work, to the papers they mention, and to my own experimental work, it is possible to avoid the objection that by making a judicious selection from the huge mass of data, good and bad, which has already been published, it is possible to prove almost anything.

2—SOLUBILITY LIMITS IN PRIMARY SOLID SOLUTIONS

An α-solid solution boundary typical of many silver and copper series is shown in fig. 2.

In some series the β-solid solution is stable down to low temperatures, and in others the solubility decreases from the point B. In the first group, therefore, only this point is a critical point. Sometimes the β-phase undergoes two changes, when there will be a point additional to B and C to be considered. In Table II the results for the point B obtained by Hume-Rothery, Mabbott, and Channel Evans, together with the other results quoted by them, are set out. It is assumed that the temperature of the solidus BD is accurately known. The point B can then be found by producing the solidus AB or the phase boundary CB until it cuts BD. In general, owing to experimental error, the two methods will not give identical results. I have re-drawn the diagrams from the data provided and have surveyed the evidence in order to find the most probable position of B. My result and that of the authors, which are often not identical, are set out in the Table II. The revised value is used in the calculations.

It was found by trial and error that the ratio of solute atoms to solvent atoms for these solid solutions, with one exception, could be put into the form 3 solute atoms with an integral number of solvent atoms. For zinc in copper, the ratio was very nearly 3 atoms to 6·5 atoms. The differences between the experimental results and those required by the stated simple ratio are expressed in Table II, in weight percentages of solute, because this form gives a better measure of the experimental error. These differences are returned to the nearest 0·1% of solute because the difficulties in work of this kind are so great that the use of a smaller unit would be almost meaningless. It will be seen that in four systems the experimental results diverge appreciably from simple ratios; but it is in the examination of these systems that very great experimental difficulties must be overcome. In my view, all divergences from simple ratios can be attributed to experimental error.

---

* 'Phil. Trans.,' A, vol. 233, p. 1 (1934).

With the copper-tin alloys an insurmountable difficulty in the consideration of those systems where the solid solubility is small is beginning to be of importance. This difficulty is simply that to show an integral numerical relationship experimental results here must needs be almost impossibly accurate. Thus, suppose the solubility of one element in a second to be 5 atomic %, giving a ratio of 1 : 19. If the measured solubility was 5·03%, the ratio would appear to be 1 : 18·9. If the three rule is true, the correct ratio is 3 : 57, and the measured ratio 3 : 56·7. This last number is too far from an integer to prove anything. For this reason all systems showing a solid solubility of less than about 10 atomic % must at present be left out of consideration. The atomic weights of the elements concerned are

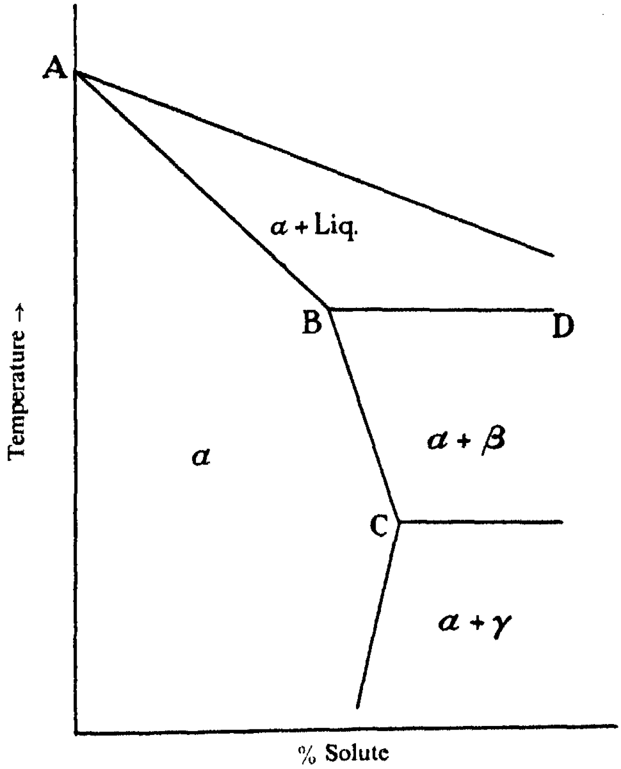


FIG. 2.

TABLE II—SILVER AND COPPER

System	Observer*	Solubility in atomic % original paper	Solubility in atomic % revised value
Ag–Cd	H.-R., M., and C. E.	37·0 & 37·3	37·0
Ag–Zn	Carpenter and Whiteley	Not stated	32·2
Cu–Zn	Genders and Bailey, H.-R., M., and C. E.	31·89 31·5	31·7
Ag–Hg	Murphy	30·56	30·4
Ag–In	H.-R., M., and C. E.	20·0	20·0
Ag–Ga	„	19·0	18·7
Ag–Al	„	Not stated	17·7
Cu–Be	Borchers	16·6	16·7
Cu–Ga	H.-R., M., and C. E.	Not stated	16·25
Cu–Al	Stockdale	15·85	15·65
Ag–Cu	Stockdale	14·07	14·07
Ag–Sn	H.-R., M., and C. E.	12·2	12·1
Cu–Ge	„	12·0	12·0
Cu–Si	Smith	11·15	11·05
Cu–Sn	Stockdale	7·59	7·55
Cu–Mg	Jones	Not stated	6·83
Cu–As	Hanson and Marryat	6·22	6·22
Cu–Sb	H.-R., M., and C. E.	Not stated	5·64
Cu–Ag	Stockdale	5·0	5·0
Ag–Be	Sloman	3·5	3·0

* The necessary references are to be found in the paper

also an important factor, and in those cases where that of the solute is much smaller than that of the solvent, as in the beryllium-copper series, results must be accepted with caution. Thus, although the work of Borchers invites confidence, it must be more or less accidental that his results fit so excellently into the theory.

Table II shows that when the systems are set out in the order of decreasing α-solubility at the B point, the alloys fall into four groups. In the first group, where the solubility is very large, the solute atoms are divalent. In the next group the solutes are all trivalent (with the exception that beryllium is occupying an anomalous position), and then come the tetravalent solutes. The solubility of copper in silver places that element apparently anomalously between the tri- and tetravalent elements. In the final group, where the α-solubility is small, the solutes may be of any valency.

SOLID SOLUTIONS, B POINTS

% solute required by suggested ratio	Difference wt. %	No. of solvent atoms dissolving 1 solute atom	Suggested simple ratio	Ratio of atomic diameter
37·50	0·5	1·703	3: 5 (1·67)	1·055
33·33	0·9	2·11	3: 6 (2·00)	1·051*
31·58	0·1	2·154	6: 13 (2·167)	1·070
30·00	0·5	2·29	3: 7 (2·33)	1·076
20·00	Nil	4·00	3: 12 (4·00)	1·090
18·75	Nil	4·35	3: 13 (4·33)	1·161*
17·65	Nil	4·65	3: 14 (4·67)	1·007*
16·67	Nil	4·99	3: 15 (5·00)	1·133*
16·67	0·45	5·15	3: 15 (5·00)	1·033*
15·79	0·1	5·39	3: 16 (5·33)	1·117
14·29	0·15	6·11	3: 18 (6·00)	1·125*
12·0	0·1	7·26	3: 22 (7·33)	1·097
12·0	Nil	7·33	3: 22 (7·33)	1·086
11·11	Nil	8·05	3: 24 (8·00)	1·094*
7·69	0·2	12·25	3: 36 (12·00)	1·235
		13·64		1·250
		15·08		1·094
		16·73		1·258
		19·00		1·125
		32·3		1·274*

by Hume-Rothery, Mabbott, and Channel Evans.

The ratios of the atomic diameters are set out in the final column of Table II, an asterisk being used to denote the systems in which the diameter of the solute is smaller than that of the solvent. These diameters are taken for co-ordination number 12 from Goldschmidt's[*] Table (Silicon for C.N.4, Gallium from Hume-Rothery).[†] There is considerable doubt about the validity of these measurements, but even if they are appreciably in error, it is difficult to find a close relationship between solubility and size. For example, if the atomic diameter of aluminium is somewhat smaller than that given by Goldschmidt, the aluminium-silver ratio as set out in Table II will be increased while that of aluminium-copper will be decreased. The mean of the two will still be near 1·06. The sum of the solubilities of aluminium in silver and in copper should therefore be about the same as that of zinc in the two solvents, provided solubility is determined solely by the size-factor. On the evidence available, therefore it is difficult to avoid the acceptance of the views of Hume-Rothery, Mabbott, and Channel Evans, which may be summarized as follows :

---

[*] 'Trans. Faraday Soc.,' vol. 25, p. 253 (1929).

[†] 'Phil. Trans.,' A, vol. 233, p. 1 (1934).

Where the atomic diameters of the solvent and solute atoms do not differ by more than a certain percentage the range of the primary solid solution is considerable and the solubility limits follow simple valency laws. Where the size-factor is unfavourable the range is very restricted.

TABLE III

	Solubility in atomic %	Suggested simple ratio	Difference in atomic numbers
Ag–Cd	37·0	3: 5	1
Ag–Zn	32·2	3: 6	17
Cu–Zn	31·7	6: 13	(1)
Ag–Hg	30·4	3: 7	33
Ag–In	20·0	3: 12	2
Ag–Ga	18·7	3: 13	16
Ag–Al	17·7	3: 14	34
Cu–Ga	16·25	3: 15	2
Cu–Al	15·65	3: 16	16
Ag–Sn	12·1	3: 22	3
Cu–Ge	12·0	3: 22	3
Cu–Si	11·05	3: 24	15
Cu–Sn	7·55	3: 36	21

If the alloys of copper with silver and with beryllium are omitted and the copper-tin alloys brought in, three clear valency groups are obtained, *see* Table III. Certain relationships can be traced within each group. For example, with the exception that the solubility of zinc in copper is greater than that of mercury in silver, it is seen that silver is always a better solvent than copper for elements of the same valency. This may not be unconnected with the fact that the atomic diameter of silver is 2·88 A, while that of copper is only 2·56 A. Further, the position of the solute in the Periodic Table with respect to that of the solvent appears to have some influence on the solubility within each valency group. The nearer the two are, the greater the solubility. The difference in the atomic numbers of solvent and solute, which is some measure of their relative positions, is given in Table III.

When we pass on to consider binary systems other than those which have been worked out by Hume-Rothery and his colleagues we are at once confronted by such a mass of conflicting evidence that the selection of data becomes a matter of great importance and must necessarily be open to criticism. For that reason I am not now going to consider this group in detail, but, if a critical examination is made of the available papers on systems which show a considerable range of solubility and in which the atomic weights of the two elements are similar, it becomes clear that there is a distinct possibility that the simple ratio rule, though not the more limited three rule, is also applicable in those alloys where the solvent is neither silver nor copper. Daeves* has already pointed out that 12 atoms of iron can dissolve 1 atom of carbon at the temperature of the eutectic and that other simple integral relationships exist in the iron-carbon system and other systems containing iron and a second element.

3—OTHER CRITICAL POINTS ON THE α-PHASE BOUNDARY

These points are of two types. In the first type the neighbouring β-phase undergoes some small change, probably a polymorphic transformation, as in the copper-zinc alloys. In the second, the β-phase breaks down completely into a mixture of α and a new phase whose crystal structure is unrelated either to that of the α or of the β. In this latter type of reaction the relationship between the phases is similar to that of the α, β, and liquid phases in a eutectic reaction. By analogy, therefore, if the saturated α-phase at the eutectic temperature contains atoms in a simple ratio, it is to be expected that at these C points, fig. 2, there would also be simple ratios. Reactions of the first type are more obscure. It will be shown below that in neither case is it probable that there is an exact simple ratio rule. Systems other than those quoted by Hume-Rothery, Mabbott, and Channel Evans will be excluded from the discussion.

Of the first fifteen systems mentioned in Table II, six (silver with mercury, indium, gallium, copper and tin, and copper with germanium) have no known C points and are not now further considered. I have made a critical examination of the data for the other systems given by Hume-Rothery, Mabbott, and Channel Evans in their paper and also of recent evidence from other sources and have concluded that in the case of five points in three systems the probable accuracy of the experimental results does not justify the use of the existing data to prove or disprove an exact simple ratio rule. For five points in three systems, the evidence is definitely against the rule; for three points in three systems it is definitely in its favour.

---

* 'Z. anorg. Chem.,' vol. 115, p. 290 (1921).

If the enquiry is limited to those binary alloys about whose constitution there is now general agreement, the results are similar. The solubility relationships in these systems, the copper-zinc, aluminium-copper, and aluminium-silver, have been investigated many times and are now probably known with great accuracy. In the case of the first, 8 atoms of copper are found to dissolve exactly 5 atoms of zinc at the temperature of the polymorphic transformation in the β-phase. The solubility of aluminium in copper was further tested experimentally. Alloys containing 9·55 and 9·60 weight % of aluminium (20 atomic % requires 9·59 weight %) were given different heat treatments. In some of the experiments the first alloy was obtained as a uniform solid solution, but the second always contained traces of another phase. There is, therefore, evidence that at the C point in this system the atoms are in the ratio 1 : 4. But in the aluminium-silver system there are two points and there is a considerable body of evidence in support of the conclusion that at both the solubility is slightly greater than that required by the 1 : 4 ratio.

The results are collected in Table IV which is set out in the same form as Table II, where data about the B points are summarized. In Table IV, points which must be regarded as evidence for an exact rule are marked with an asterisk, those against with a dagger, and those points about which the evidence is inconclusive are left unmarked.

It will be seen that there appears to be no connexion between the ratios, whether integral or not, and the nature of the transformation in the β-phase, and also that though no exact rule can be established there is some possibility that it is the determining factor in solid solutions of this kind, but that it is being obscured by some minor disturbing factor.

Table IV also shows that the solubility at the C points follows closely that at the B points, bivalent solutes being generally the most soluble, trivalent less, and tetravalent least. It so happens that in many alloys the solubility at a C point is also the maximum solubility, and therefore calculation must necessarily show that the ratios given in the penultimate column of this table are in general conformity with the Hume-Rothery rule, " Where the atomic size-factors are favourable, the maximum α-solid solubility, to a first approximation, corresponds to a constant electron concentration." It is noteworthy that for trivalent solutes the atom-electron ratio is always almost exactly 5 : 7.

There also appears to be an approximation to a simple relationship between the dissolving powers of a solvent for any one solute at the B and C points. The dissolving power of a solvent may be defined as being

TABLE IV—SILVER AND COPPER SOLID SOLUTIONS, C POINTS

System	Solubility in atomic % original paper	Solubility in atomic % revised value	% solute required by ratio	Difference weight %	No. of solvent atoms dissolving 1 solute atom	Nearest simple ratio	Nature of β-transformation
Ag–Cd (1)	42·4	—	42·86	0·5	1·36	3: 4 (1·33)	$\beta \rightarrow \beta'$
(2)	Near 42·0	—	42·10	0·1	1·38	8: 11 (1·38)	$\beta \rightarrow \beta''$
Ag–Zn*	37·3	—	37·5	0·2	1·68	3: 5 (1·67)	$\beta \rightarrow \alpha + \gamma$
Cu–Zn*	Not stated	38·43	38·46	Nil	1·60	5: 8 (1·60)	$\beta \rightarrow \beta'$
Ag–Al (1)†	20·65	20·4	20·0	0·15	3·90	1: 4	$\beta \rightarrow \alpha + \gamma$
(2)†	—	20·3	20·0	0·10	3·93	1: 4	$\alpha + \gamma \rightarrow \beta'$
Cu–Be	9·1	—	9·09	Nil	10·00	1: 10	$\beta \rightarrow \alpha + \gamma$
Cu–Ga†	20·25	20·3	20·0	0·3	3·93	1: 4	Unknown
Cu–Al*	20·39	Near 20·0	20·0	Nil	4·00	1: 4	$\beta \rightarrow \alpha + \gamma$
Cu–Si (1)†	14·0	—	14·29	0·15	6·14	1: 6	Unknown
(2)†	14·0	—	14·29	0·15	6·14	1: 6	Unknown
Cu–Sn (1)	Not stated	8·90	9·09	0·3	10·24	1: 10	$\beta \rightarrow \beta'$
(2)	9·26	9·27	9·09	0·3	9·79	1: 10	$\beta' \rightarrow \alpha + \gamma$

inversely proportional to the number of solvent atoms required to dissolve one solute atom. Thus in the copper-aluminium series the dissolving power at the B point is very nearly ¾ that at the C point, and in the copper-beryllium alloys it is twice that at the C point.

An empirical rule for calculating the solubility at the C point can be established from the available experimental data in the case of those systems where the solubility increases as the temperature falls, provided the solubility at the temperature of the peritectic horizontal is known. This rule can be stated as follows: If the dissolving power of the solvent at the B point with respect to a solute be expressed as being inversely proportional to a small whole number, N, where N is 4, 5, 6, or 7, and is such that its product with an integer, A, is the number of atoms (taken to the nearest whole number) dissolving 3 atoms of solute, then the dissolving power at the C point is inversely proportional to N − 1, and the number of solvent atoms dissolving 3 solute atoms at the C point is approximately (N − 1) A.

That is—
$$\frac{\text{Dissolving Power at B point}}{\text{Dissolving Power at C point}} = \frac{N-1}{N}.$$

TABLE V

System	Nearest simple ratio at B point	Nearest simple ratio at C point	N	A	Solubility at C point observed	Solubility at C point calculated
Ag–Cd	3: 5	(1) 3: 4	5	1	42·4	42·3
		(2) 8: 11	—	—	—	—
Ag–Zn	3: 6	3: 5	6	1	37·3	36·3
Cu–Zn	6: 13	5: 8	—	—	38·4	38·2
Ag–Al	3: 14	3: 12	7	2	20·4	20·1
Cu–Be	3: 15	3: 30	—	—	—	—
Cu–Ga	3: 15	3: 12	5	3	20·3	19·5
Cu–Al	3: 16	3: 12	4	4	20·0	19·8
Cu–Si	3: 24	3: 18	4	6	14·0	14·2
Cu–Sn	3: 36	(1) 3: 30	6	6	9·27	8·92
		(2) 3: 30	6	6	8·90	

The observed and calculated values for the C points obtained by the use of this formula are set out in Table V, the B point dissolving power used in making up this table being the observed and not the adjusted value. The agreement, except for the silver-zinc and copper-gallium alloys, is good, but, as has been already indicated, the experimental examination of these alloys is exceptionally difficult. The copper-zinc ratios have hitherto been expressed in such a form that no calculation can be made, but even here the dissolving powers are almost simply related. Their ratio is $1\frac{1}{4}\frac{7}{8}$ and not $1\frac{1}{4}\frac{0}{8}$, or $\frac{4}{8}$. If the solubility of zinc in copper at the B point is 31·92% (the experimental value is 31·7), giving a ratio of 30: 64 instead of 30: 65, the relationship becomes $\frac{4}{8}$, and the ratios are in a form not altogether dissimilar from the others in the series, but speculations of this kind are profitless at this stage.

4—EUTECTICS IN BINARY METALLIC SYSTEMS

When we come to consider the more complicated case of a binary system containing two phases, each of which may contain the two kinds of atoms in a simple ratio, we must expect more complicated results. It is as if we had two walls of the same area but of different thickness and built of blue and red bricks, the red bricks being larger than the blue. If the first wall contained 1 red and 8 blue bricks per unit volume and was twice as thick as the second which contained 12 red and 1 blue per unit volume, the ratio of bricks in the two walls would be 18: 13. However, we can also consider the ratio between the bricks of different colour. Here the ratio would be 14 red to 17 blue. That this simple analogy leads to more complicated ratios than many of those which have been determined experimentally is remarkable.

As a first step towards proving or disproving the statement that in a binary eutectic the atoms of the two metals are present in a simple integral ratio the compositions of six eutectic mixtures, four between two primary solid solutions and two between a primary and a secondary solid solution, were investigated.* The method used was to determine the freezing-points of a series of alloys of almost the eutectic composition and to find the eutectic point by extrapolations of the liquidus lines. As the alloys of each system were examined under similar experimental conditions, it was possible to check the results by plotting the times elapsing between the first appearance of pre-eutectic metal and the first appearance of the eutectic mixture, against composition. The results are shown in Table VI, together with W. R. D. Jones's determination of the composition of the $Mg - Mg_2Cu$ eutectic.† Jones's appears to be the only recent examination of an alloy system of this type which has been carried out with the necessary accuracy, and in which the eutectic mixture contains considerable percentages of both metals.

---

* 'J. Inst. Met.,' vol. 43, p. 193 (1930).

† 'J. Inst. Met.,' vol. 46, p. 395 (1931).

As a basis for the discussion of these results, a simple ratio was quite arbitrarily defined as a ratio between any integral number up to 8 atoms of one constituent and any integral number up to 24 of the other. This definition brings in a large number of ratios. For example, in the range 25·00 to 33·33 atomic percentages there are 22. In the absence of the method of plotting given later in this paper it was necessary in weighing the evidence to consider the average interval between each ratio. In the range mentioned above, this is only 0·38 atomic percentages. Therefore, if there is a simple ratio rule, the experimental results given in Table VI, considered in terms of the average interval, are not of an accuracy great enough to prove it. It was suggested in the earlier paper that the balance of evidence was against the rule, but, as will be shown later, the new method of graphical analysis indicates that there is a definite probability of its existence.

TABLE VI

System	Eutectic composition by experiment atomic %		Atomic ratio	Nearest integral ratio	Atomic % required by integral ratio	Difference by weight %
Ag–Cu	39·83 Cu	60·17 Ag	1 : 1·511	2 : 3 (1·50)	40·00 Cu	0·14
Cd–Sn	33·45 Cd	66·55 Sn	1 : 1·989	1 : 2 (2·00)	33·33 Cd	0·12
Cd–Zn	26·59 Zn	73·41 Cd	1 : 2·761	4 : 11 (2·75)	26·67 Zn	0·06
Pb–Sn	26·10 Pb	73·90 Sn	1 : 2·831	6 : 17 (2·833)	26·09 Pb	0·01
Ag–Sb	41·05 Sb	58·95 Ag	1 : 1·436	7 : 10 (1·429)	41·18 Sb	0·13
Al–Cu	17·37 Cu	82·63 Al	1 : 4·757	4 : 19 (4·75)	17·39 Cu	0·02
Cu–Mg	14·52 Cu	85·48 Mg	1 : 5·887	8 : 47 (5·875)	14·55 Cu	0·05

In a discussion, Hargreaves* pointed out that integral atomic ratios between constituents were not to be expected, but that there might be simple ratios between the number of atoms in the two phases. This possibility had already occurred to me.

A typical eutectic diagram is given in fig. 3.

If the ratio is between the constituents, then the lengths AC and CE are simply related. If Hargreaves's suggestion is correct, BC and CD are simply related. To examine this it was necessary to determine the position of the points B, C, and D.† The results are collected in Table VII. All percentages in this table, except those given in the last column, are atomic and the lettering is in accordance with that of fig. 3.

---

* 'J. Inst. Met.,' vol. 43, p. 413 (1930).

† The results for five of the seven systems of Table VII have already been published (*see* footnote, p. 85). The full results for the antimony-silver and cadmium-tin alloys have not appeared in print. The necessary solidus curves of these two systems have been determined by micrographic methods similar to those used in the investigation of the aluminium-rich aluminium-copper alloys (Stockdale, 'J. Inst. Met.,'

Errors in the measurements of the B and D points tend to cancel each other in that if the solubility of the first metal in the second metal of a system is over-estimated so also will that of the second in the first probably be over-estimated and vice versa. Further, errors in the measurement of B or D will affect the lengths of BC or CD only, while an error in the measurement of C will affect the lengths of both lines. Therefore it is assumed for the purpose of calculating the differences shown in the last column of Table VII that the positions of the B and D points are known accurately and that any divergences are due to faulty measurements of the position of the C points.

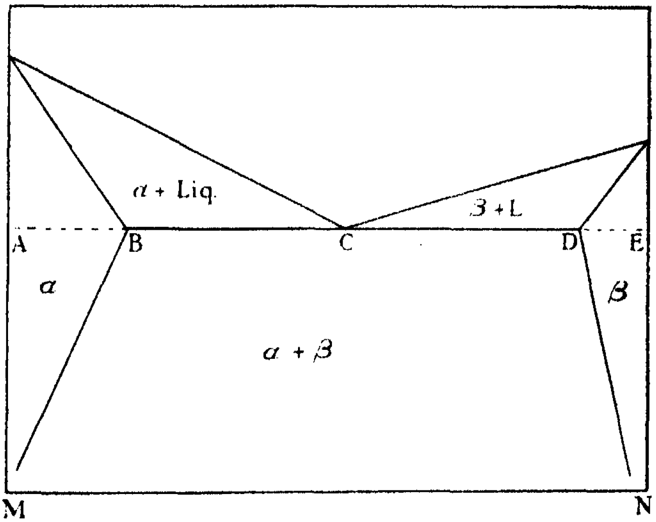


FIG. 3.

Table VII shows at once that for the first five systems there is a very close approximation to a simple integral ratio between the numbers of atoms in each phase. (This kind of ratio will henceforward be referred to as the phase ratio, and the other will be known as the constituent ratio.) For the other two systems the agreement is not so close, but in both these systems

vol. 52, p. 111 (1933)). The solubility of silver in antimony was found to be considerably less than 0·1% and may be neglected. The solidus curve for "Ag_3Sb" closely resembles that for "$CuAl_2$." The solubility at the temperature of the eutectic was found to be 27·25% by weight of antimony (Ag_3Sb requires 27·34) and it appeared to be unchanged down to 400° C. Experiments with the other alloys showed that at the temperature of the eutectic the solubility of tin in cadmium was 0·25% and of cadmium in tin 5·5% by weight.

TABLE VII

System	Composition of eutectic % N C	Solubility of N in M % N B	100-solubility of M in N % N D	Approximate composition of secondary solid solution	$\frac{BC}{CD}$	Nearest integral ratio: atoms M rich phase: atoms N phase	Composition of eutectic required for integral ratio % N	Difference required—observed % by weight
M N								
Cu–Ag	60·17	5·00	85·93	—	$\frac{55\cdot17}{25\cdot76} = 2\cdot142$	7: 15 (2·143)	60·18	0·01
Cd–Sn	66·55	0·24	94·21	—	$\frac{66\cdot31}{27\cdot66} = 2\cdot397$	5: 12 (2·400)	66·57	0·02
Zn–Cd	73·41	1·26	95·04	—	$\frac{72\cdot15}{21\cdot63} = 3\cdot336$	3: 10 (3·333)	73·40	0·01
Pb–Sn	73·90	29·71	98·50	—	$\frac{44\cdot19}{24\cdot60} = 1\cdot796$	5: 9 (1·800)	73·93	0·04
Sb–Ag	58·95	Nil	75·08	Ag_3Sb	$\frac{58\cdot95}{16\cdot13} = 3\cdot656$	3: 11 (3·667)	58·99	0·04
Cu–Al	82·63	68·10	97·49	$CuAl_2$	$\frac{14\cdot53}{14\cdot86} = 0\cdot977$ or 1·023 : 1	1: 1 (1·000)	82·80	0·26
Cu–Mg	85·48	66·67	99·99	$CuMg_2$	$\frac{18\cdot81}{14\cdot51} = 1\cdot296$	7: 9 (1·286)	85·41	0·12

there is a reactive element present, and the further complication that the alloys have to be protected very carefully from contamination when they are molten adds greatly to the experimental difficulties in the determination of the various points. Further, the ratios of the atomic weights of the elements are unfavourable and the lengths of both BC and CD are small. In my opinion, even in the aluminium-copper system the divergence from an exact simple ratio can be attributed to experimental error.

Let us define a simple ratio as any integral number of atoms up to 8 of one element with any integral number of atoms of the second. Consider the series 8 atoms with *n* atoms. The ratios are 8 : 8, 8 : 9, 8 : 10 —— 8 : 8 + 8, 8 : 8 + 9, 8 : 8 + 10 ——, or 1 : 1·0, 1·125, 1·25 —— 2·0, 2·125, 2·25 —— 3·0, 3·125, 3·25, etc. For the series with 7 atoms they are 1 : 1·0, 1·143, 1·286 —— 2·0, 2·143, 2·286, etc.

According to this definition, between the ratios 1 : 1 and 1 : 2, there are 22 simple ratios, and between 1 : 2 and 1 : 3 there are also 22, etc. The ratio 7 : 9, or 1 : 1·286 in one range is followed by 7 : 7 + 9 or 1 : 2·286 in the next, and by 1 : 3·286 in the next, and so on. For this reason, in considering the experimental results, it is unnecessary to consider the units to the left of the decimal point, and a simple graphical representation becomes possible as follows:

Draw a line LMM′NN′—— Z one thousand units in length, fig. 4, and mark on it the ratios as defined above, neglecting the units. Thus the ratios 4 : 5 (1 : 1·25) and 4 : 9 (1 : 2·25) are both represented on the line by the point A at 250. The next point to be marked, B at 286, represents the ratios 7 : 9, 7 : 16, etc. The length of AB is then 36 units. Divide this into three so that 2AC = 2DB = CD. Proceed in this way for all the other points. Mark in the lengths corresponding to AC and DB, leaving the rest of the line blank. Then the sum of the lengths of the marked portions is equal to that of the blank portions, and if there is no simple ratio rule the experimental ratios should fall at random on the marked and on the unmarked. Fig. 4 shows that this is by no means the case, and that there is a definite probability of the existence of a simple ratio rule.

Of the seven ratios, only one, the antimony-silver, falls in an unmarked space; but Table VII shows that this result is unimportant because the large proportion of antimony atoms in the silver-rich phase calls for experimental work of the highest accuracy. On the other hand, fig. 4 disguises the erratic result for the aluminium-copper ratio. A calculation of the exact odds in favour of the existence of the rule on the lines suggested by this diagram is therefore undesirable.

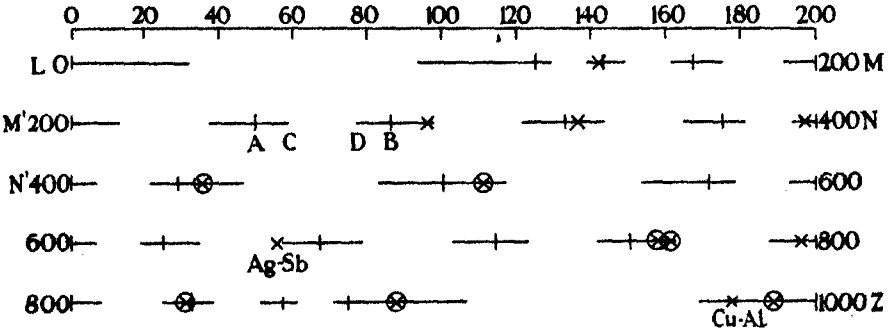
If the constituent ratios set out in Table VI are also plotted in fig. 4, the curious result is obtained that they also appear to conform to a simple ratio rule. Indeed, the evidence for such a rule appears to be more definite in this case than in the case of the corrected ratios.

In this paper, then, evidence in favour of the following three rules is put forward:

(i) The solute and solvent atoms are present in a simple integral ratio in a saturated solid solution at the temperature of the eutectic or peritectic horizontal.

(ii) In a eutectic mixture the elements are present in a simple integral atomic ratio.

(iii) In a eutectic mixture the ratio of the number of atoms, irrespective of their kind, in the phases is simple.



× Phase ratios (Table VII). ⊗ Constituent ratios (Table VI).

FIG. 4.

It can be shown by calculation that for these seven systems not more than two of the above rules can be exactly true. As an example, consider the copper-silver system. Here the phase ratio is almost exactly 7 atoms of copper-rich phase with 15 atoms of silver-rich phase. By experiment, one atom of silver dissolves in exactly 19 atoms of copper and one atom of copper in nearly 6 atoms of silver. The determined solubility of copper in silver is 14·07 atomic percentages while the ratio requires 14·29 percentages. Assume that rules (i) and (iii) are true. The ratios are then respectively 7/15, 1/19, 1/6. Consider a eutectic alloy containing 7 × 20 × 7 atoms in the copper-rich phase and 15 × 20 × 7 atoms in the silver-rich phase. In the former there are 7 × (20 − 1) × 7 atoms of copper and 7 × 7 of silver; in the latter 15 × 20 (7 − 1) atoms of silver and 15 × 20 atoms of copper. That is, there are present 1231 atoms of copper and 1849 atoms of silver, a ratio so nearly 2:3 that the difference from that ratio cannot be determined by the present experimental methods. The other systems work out in a similar way, with very small differences between the determined and calculated constituent ratios, although, in certain cases where the solubility of one metal in the second is small, it is necessary to make somewhat bold assumptions about the solubility ratios. It is not clear whether this result is accidental or whether it has some physical meaning, but quite definitely it is arithmetically impossible for all the rules to be exact. If (i) be accepted, then either (ii) or (iii) must be false. There is not enough evidence available to enable a decision to be made between them, but I am of the opinion that (iii) is true, because in those systems where the eutectics are between primary solid solutions and which are easiest to investigate experimentally the experimental results conform more nearly to a simple phase ratio, Table VII. The mean difference for these four systems is 0·02% by weight as against 0·08% for the ratios between the numbers of atoms of each kind, Table VI. In the other systems, although the mean difference is 0·07% against 0·14% in favour of the constituent ratios, yet the fact that the phase ratios are the simpler is probably evidence in their favour.

The figures given in the third last column of Table VIII show how closely the calculated constituent ratios conform to rule (ii) and those in the last two columns compare the observed compositions for the various eutectics with the compositions calculated from the adjusted experimental results, on the assumptions that rules (i) and (iii) are exact and that the experimental determinations have been sufficiently accurate to indicate the correct ratios.

There appears to be no *a priori* reason why these rules should not be applicable to non-metallic systems. Calculations have been made for the following types of eutectic series: salt-water, binary organic compounds, fused salts. The evidence in favour of the existence of simple ratios in cryohydrates is not good. This may be due to the complications introduced by ionization or to the probability that the experimental work has not been of an accuracy great enough for this purpose. The mere fact that the molecular weight of water is so small reduces the permissible error in the determination of the eutectic compositions in series of this type to an almost negligible figure. There is some evidence that in eutectics containing two organic compounds the molecules are present in a simple ratio, but, as the attention of organic chemists is most frequently directed to the formation of compounds in binary systems, the information about the composition of these eutectics is neither so extensive nor so exact as is desirable. Binary salt systems, however, do seem to obey the simple ratio constituent rules. For example, the work of Madgin and his colleagues* on fused nitrates has been examined and ratios similar in character to those set out in Table VI have been obtained.

* Laybourn, Madgin, and Freeman, 'J. Chem. Soc.,' p. 139 (1934), *inter alia*.

TABLE VIII

System	Solubility ratios	Solubility required by ratio Atomic %	Solubility observed atomic %	Phase ratio at eutectic Table VII	Suggested constituent ratio at eutectic Table VI	Constituent ratio calculated	Difference calculated and suggested Atomic %	Eutectic composition % by weight	
								Calculated	Observed
Cu–Ag	1 Ag 19 Cu 1 Cu 6 Ag	5·00 14·29	5·00 14·07	7/15	$2/3 = \frac{3698}{5547}$	$\frac{3693}{5547}$	$\frac{5}{5547} \equiv 0{\cdot}03$	28·18 Cu	28·06 Cu
Cd–Sn	1 Cd 16 Sn ?	5·88 Nil	5·79 0·24	5/12	$1/2 = \frac{96}{192}$	$\frac{97}{192}$	$\frac{1}{192} \equiv 0{\cdot}23$	32·36 Cd	32·25 Cd
Zn–Cd	1 Zn 19 Cd 1 Cd 79 Zn	5·00 1·25	4·97 1·26	3/10	$4/11 = \frac{3052}{8393}$	$\frac{3047}{8393}$	$\frac{5}{8393} \equiv 0{\cdot}04$	17·43 Zn	17·40 Zn
Pb–Sn	3 Sn 7 Pb 1 Pb 63 Sn	30·0 1·56	29·7 1·50	5/9	$6/17 = \frac{234}{663}$	$\frac{233}{663}$	$\frac{1}{663} \equiv 0{\cdot}09$	38·02 Pb	38·14 Pb
Sb–Ag	1 Sb 3 Ag —	75·00 Nil	75·08 Nil	3/11	$7/10 = \frac{231}{330}$	$\frac{230}{330}$	$\frac{1}{330} \equiv 0{\cdot}11$	44·03 Sb	44·00 Sb
Al–Cu	8 Cu 17 Al 1 Cu 39 Al	68·00 2·50	68·10 2·52	1/1	$4/19 = \frac{1324}{6289}$	$\frac{1311}{6289}$	$\frac{13}{6289} \equiv 0{\cdot}14$	32·95 Cu	33·15 Cu
Cu–Mg	1 Cu 2 Mg —	33·33 Nil	33·33 0·01	7/9	$8/47 = \frac{328}{1927}$	$\frac{329}{1927}$	$\frac{1}{1927} \equiv 0{\cdot}04$	30·85 Cu	30·75 Cu

5—CONCLUSION

The possibility that simple integral relationships are of the greatest importance in the determination of solubility and other limits in the equilibrium diagrams of binary systems is pointed out. The rules now suggested can only be established or disproved when a sufficient number of accurate experimental results has been accumulated. Only α-solid solutions and eutectics, the two groups which are probably the easiest to examine experimentally and where small experimental errors least obscure the results, have now been discussed, but it may be that as experimental technique improves it will be possible to formulate other rules to cover other groups of binary and even ternary systems.

The production of the Tables given in this paper has involved the making of hundreds of small calculations. These have been checked and rechecked and it is hoped that the paper is now free from arithmetical error. The use of C. S. Smith's Atomic and Weight Percentage Interconversion Tables* has very greatly facilitated this side of the work.

I wish to thank Professor Sir William Pope for communicating this paper and Professor R. S. Hutton and Professor J. E. Lennard-Jones for their criticism.

6—SUMMARY

The Hume-Rothery valency rules are summarized and the work of Hume-Rothery, Mabbott, and Channel Evans on the depressions of the melting- and freezing-points of copper and silver by the addition of elements of the B sub-groups, together with the solid solubility limits in these alloys, is discussed. The calculations given in this paper show that it is probable that there are other integral relationships between the numbers of atoms in a binary alloy. The following empirical rules are put forward:

(*a*) In a saturated solid solution at the temperature of the eutectic or peritectic horizontal there is a simple integral relationship between the numbers of solvent and solute atoms.

(*b*) There is a similar relationship in a saturated solid solution in contact with a second solid phase at the temperature of transformation of the second phase.

(*c*) In copper and silver-rich alloys, the solubilities at the above two temperatures are simply related and it is possible to predict the solubility at the lower temperature if that at the higher is known.

---

* 'Trans. Amer. Inst. Min. and Met. Eng. Iron and Steel and Inst. Met. Divisions,' Contrib. No. 60 (1934).

(*d*) In a eutectic mixture the elements are present in a simple integral atomic ratio.

(*e*) In a eutectic mixture the ratio of the numbers of atoms, irrespective of their kind, in the two phases is simple.

It may be that none of these five rules is exact, but their use gives results which are more accurate than the majority of those which have been obtained experimentally. In particular, recent experimental work shows that rule (*b*) leads to an under-estimation by about 1 part in 40 of the solubility of certain metals in copper and silver. It is also improbable that rules (*d*) and (*e*) are both exact, but if either is true, then, in the systems investigated, the other is true to within at least 1 part in 200.

Diffraction of Electrons by Metal Crystals and by Mica

By J. A. DARBYSHIRE, M.Sc., Ph.D., and E. R. COOPER, M.Sc. (1851 Exhibition Scholar, New Zealand)

(*Communicated by W. L. Bragg, F.R.S.—Received March* 23—*Revised May* 31, 1935)

[PLATES 3–4]

INTRODUCTION

The object of the research was to study the spot patterns produced when an electron beam of homogeneous velocity ~ 30 kV is transmitted through a thin crystalline film. Such spot patterns were first reported by Kikuchi,* and their importance lies in the fact that they are regarded as being due to single crystals of matter; the study of the diffraction of electrons by single crystals must necessarily precede the exact explanation of the Debye-Scherrer patterns obtained when random aggregates are employed. The results described below have a direct bearing on the latter problem.

Electron diffraction spot patterns due to the transmission of 30 kV electrons have been studied by Thomson,† Kirchner,‡ Trillat and Hirsch,§

* 'Jap. J. Phys.,' vol. 5, p. 83 (1928).

† 'Proc. Roy. Soc.,' A, vol. 133, p. 1 (1931).

‡ 'Phys. Z.,' vol. 32, p. 969 (1931).

§ 'Z. Physik,' vol. 75, p. 784 (1932).

Lassen.* But the interpretation of these pseudo-two-dimensional effects has remained uncertain.

This uncertainty has arisen from the fact that many diffracted electron-beams appear simultaneously, even though the electron-beam incident on the diffracting material is of homogeneous velocity. If the diffraction of electrons by a single perfect crystal took place strictly according to the equations

$$\lambda = 2d_{hkl} \sin \theta = h \frac{\sqrt{1 - \beta^2}}{m_0 v}$$

(where the symbols have their usual significance), then only one diffracted beam could appear at a time. Four possible explanations have been advanced to account for the simultaneous appearance of so many diffracted beams, each of which will be described more fully in the next section, viz., A, low resolving power due to small crystal size. This was suggested by Bragg and Kirchner† and is generally referred to as the Bragg-Kirchner effect. B, Distortion of the diffracting crystal. This possibility was put forward by Bragg‡ soon after the experiments of Kikuchi had been carried out. C, Dynamical interaction among the secondary diffracted wavelets set up in the crystal. The dynamical theory of electron diffraction has been treated by Bethe§ and Morse.‖ D, Inelastic collisions between the electrons and the atoms of the diffracting crystal.

In what follows we show that of these factors distortion is the main cause of the apparent two-dimensional cross-grating patterns observed in our electron diffraction experiments. Results are obtained which show that interaction is of importance in determining the intensities of the diffracted electron beams.

In a previous paper,¶ the writers have described the ring patterns obtained from oxide films removed from the surface of molten metals in air. The present work consists of an examination of the spot patterns which the oxide films sometimes yielded; these spot patterns have been found to be due to pure metal crystals caught up on the oxide films, and precise measurements have been possible by using the oxide rings with which to calibrate the spot patterns.

We also include in this paper the results of experiments which have been carried out on the transmission of 30 kV electrons through thin bent sheets of mica.

---

* 'Ann. Physik,' vol. 22, p. 65 (1935).

† 'Nature,' vol. 127, p. 738 (1931).

‡ 'Nature,' vol. 124, p. 125 (1929).

§ 'Ann. Physik,' vol. 87, p. 55 (1928).

‖ 'Phys. Rev.,' vol. 35, p. 1310 (1930).

¶ 'Trans. Faraday Soc.,' vol. 30, p. 1038 (1934).

THEORY USED IN DISCUSSING RESULTS

Throughout the discussion of results, reference will be made to the reciprocal lattice method of picturing X-ray and electron diffraction due to Ewald. A detailed treatment of this method has been given by Bernal,† and reference should be made to his paper for further information. Briefly the construction is as follows.

The reciprocal lattice is derived from the crystal lattice by taking an origin O, and drawing lines OP_{hkl} normal to all possible *hkl* planes of the crystal lattice; the lengths OP_{hkl} of these lines are made equal to the reciprocals of the spacings d_{hkl} of the corresponding crystal planes. The reciprocal lattice consists, therefore, of a lattice of points P_{hkl}; these points are referred to as representative points. From the origin O, one draws a line $OC = 1/\lambda$ towards the source of radiation, and with centre C a sphere of radius $1/\lambda$ is described. This sphere is termed the "Ausbreitungskugel" by Ewald, and the "sphere of reflection" by Bernal.

If now the *hkl* crystal planes are in a position to reflect according to the Bragg law $\lambda = 2d_{hkl} \sin \theta$, then it is easily shown that the representative point P_{hkl} must lie exactly on the sphere of reflection. The diffracted beam is given by CP_{hkl}. Since, for 30 kV electrons, $1/\lambda$ is of the order 100 times the primitive translation in the reciprocal lattice, then the sphere of reflection is, for our purpose, almost a plane.

We now proceed to show how the explanations referred to in the introduction may be dealt with by means of the reciprocal lattice.

A—*Low Resolving Power due to Small Crystal Size*—There is a perfectly general theorem which states that if the diffracting crystal be of such a size that there are N_a, N_b, N_c lattice points parallel to the *a*, *b*, *c*, primitive translations of the crystal lattice, then each representative point P_{hkl} in the reciprocal lattice must be replaced by an elementary parallelopiped whose centre is at P_{hkl}, and whose sides are of lengths $2a^*/N_a$, $2b^*/N_b$, $2c^*/N_c$, measured parallel to the axes of the reciprocal lattice. On this account, therefore, it may be possible for the sphere of reflection to intersect several representative points simultaneously, so that several diffracted beams will occur together. This is essentially the same result as that arrived at by Bragg and Kirchner; for if, as an example, one considers an electron beam transmitted normally through a thin sheet of a single crystal cut normal to the *c* axis, then each representative point is to be replaced by an element of a straight line of length $2/t$ drawn parallel to OC*, where t is the thickness of the sheet. It follows directly from this construction that the angular range ϕ_0 over which diffracted beams will occur is given by

---

† 'Proc. Roy. Soc.,' A, vol. 113, p. 117 (1926).

$$\phi_0 = 2\sqrt{\frac{2\lambda}{t}},$$

which is the formula deduced by Bragg and Kirchner.

The construction employed here enables one to predict a further result due to low resolving power of the diffracting crystal. If the sphere of reflection passes through $P_{h_1k_1l_1}$ then a diffracted beam $CP_{h_1k_1l_1}$ arises. But the beam $CP_{h_1k_1l_1}$ may now be regarded as a beam incident on the crystal, and it may be reflected by a second set of planes $h_2k_2l_2$. The probability of this occurring when the crystal is large is very remote. The effect may be appreciable, however, when the electron beam proceeds normally to a plane in the reciprocal lattice containing representative points which are extended normally to the plane on account of low resolving power. It is easily shown in the latter case that the result of double reflection by the planes $h_1k_1l_1$; $h_2k_2l_2$ (where $P_{h_1k_1l_1}$, $P_{h_2k_2l_2}$ lie in the plane normal to the beam) in the single crystal is to give rise to the spectrum $h_1 + h_2$, $k_1 + k_2$, $l_1 + l_2$, and this may or may not be forbidden by structure factor conditions. We include this effect under the general heading of dynamical interaction.

B—*Distortion*—A distorted crystal may be supposed to be made up of a number of crystallites slightly inclined to one another, the relative inclinations depending upon the nature of the distortion. Then the reciprocal lattices for all these crystallites may be considered to have a common origin, but also to be inclined to one another. Alternatively, however, one can choose the reciprocal lattice for a particular crystallite of the aggregate and extend the representative points in such a manner as to describe the positions of all the remaining crystallites; the representative points extended thus will be elements of spherical surfaces, and the shapes of these elements will depend upon the nature of the distortion, and will also be related to one another.

It is otherwise obvious that a distorted crystal can give rise to several diffracted beams simultaneously, but the consideration of distortion by means of the reciprocal lattice is an advantage in electron diffraction, since the diffraction pattern may be regarded approximately as being a plane section passing through the origin of the reciprocal lattice and the extended representative points. By obtaining a sufficient number of patterns from the one crystal, the nature of the distortion may be uniquely determined.

C—*Dynamical Interaction*—The dynamical theory of electron diffraction treated by Bethe corresponds closely to the dynamical theory of X-ray diffraction developed by Ewald. The effect of this interaction has received a much simpler treatment for the case of X-rays by Darwin.* Its importance in electron diffraction, however, is greater than for X-rays, because of the larger scattering power of atoms for electrons, few crystal planes being necessary to produce a strong effect. Some evidence of the efficiency with which electrons are scattered elastically has been obtained in the experiments described below.

From Bethe's theory it appears that a diffracted beam due to the *hkl* planes may appear for a large crystal, even though the representative point P_{hkl} does not lie exactly on the sphere of reflection. The numerical data given by Bethe indicates that the effect is very small for 30 kV electrons; the theory, however, is an approximation, and it is still possible that it really has some significance for other than slow electrons (< 10 kV). The difficulties of such a direct attack on the problem are necessarily complicated.

Another result arrived at by Bethe is that representative points in the reciprocal lattice which have zero " weight " on account of zero structure factor, may gain weight on account of dynamical interaction, and so forbidden spectra may arise. These spectra will always correspond approximately to possible values of d_{hkl} in the crystal lattice. So-called " half-order " spectra cannot arise by this mechanism, and their existence is doubtful, even experimentally. In the experimental results described below, forbidden spectra are encountered, but we find that they may also be explained by distortion of the diffracting crystal, assuming successive reflection by two different crystallites.

It is necessary to understand that the dynamical theory predicts a cross-grating pattern from a single crystal; each spectrum of the pattern appears in a direction obtained, approximately, by joining the centre of the sphere of reflection to the representative point in question, and those spectra will be most intense whose representative points lie nearest the sphere of reflection.

D—*Inelastic Collisions*—The existence of black and white Kikuchi lines in the electron diffraction patterns, obtained by transmitting a beam normally through mica sheets of thickness $\sim 10^{-5}$ cm, shows that an appreciable number of inelastic collisions with very small change in energy occurs in this thickness of material. The mean free path of electrons in solid matter has been given as $\sim 4 \times 10^{-6}$ cm by Thomson, and, after traversing a distance of this order in the crystal, the majority of electrons will have suffered one inelastic collision. Such collisions will involve phase changes, so that, among the electrons of the incident beam coherent, interference may take place only over a thickness of the crystal of the order 4×10^{-6} cm. Briefly, therefore, the effect of inelastic collisions is to reduce the effective thickness of the diffracting crystal in the direction of the incident beam, and this we have seen can be represented by extending the representative points in the reciprocal lattice parallel to the vector OC.

---

* 'Phil. Mag.,' vol. 27, p. 315 (1914).

APPARATUS AND EXPERIMENTAL PROCEDURE

The apparatus employed in the present research has been described elsewhere.* By a special device it is possible to mount 20 specimens in the diffraction chamber at a time and to make a rapid survey of their diffraction patterns by visual observation on the fluorescent screen. Those specimens exhibiting spot patterns were then selected, and each was mounted separately on a special holder which enabled the specimen to be more accurately adjusted in the path of the electron beam. The apparatus was so designed that the diffracting crystal could be rotated in a plane normal to the beam, and in incidence; the rotation in incidence could be made about either of two axes at right angles, one of which was fixed perpendicular to the electron beam. This rotation apparatus was equivalent to a goniometer head.

Preparation of Specimens—Films of oxide of the metals zinc, cadmium, bismuth, and aluminium were prepared by removing the surface skin from the molten metal by means of a loop of copper wire; the diameter of the loop was 0·5 cm. The preparation was carried out in an atmosphere having an excess of nitrogen, so that oxidation was retarded, and the temperature of the metal was held just above the melting-point. Debye-Scherrer patterns due to the oxide were always obtained and occasionally spot patterns due to metal crystals caught up on the oxide film.

RESULTS AND DISCUSSION

Cadmium

Crystallographic Data—Cadmium has a structure resembling the hexagonal close-packed arrangement of atoms. Referred to hexagonal axes, the dimensions of the structure are $a = 2{\cdot}952$ A, $c = 5{\cdot}635$ A. The primitive translations of the corresponding reciprocal lattice are $a^* = 0{\cdot}391 \times 10^8$ cm and $c^* = 0{\cdot}177 \times 10^8$ cm. A perspective drawing of this reciprocal lattice is given in fig. 1, all planes *hkjl* being represented, whatever their structure factors. The structure factor F_{hkl} for Cd is given by

---

* Darbyshire and Cooper, 'J. Sci. Instr.,' vol. 12, p. 10 (1935).

$$F^2_{hkl} = 2f_{\text{Cd}}^2 \{1 + \cos \pi/3\,(2h + 4k + 3l)\},$$

so that spectra for which $2(h + 2k) + 3l$ is an odd multiple for 3 are forbidden.

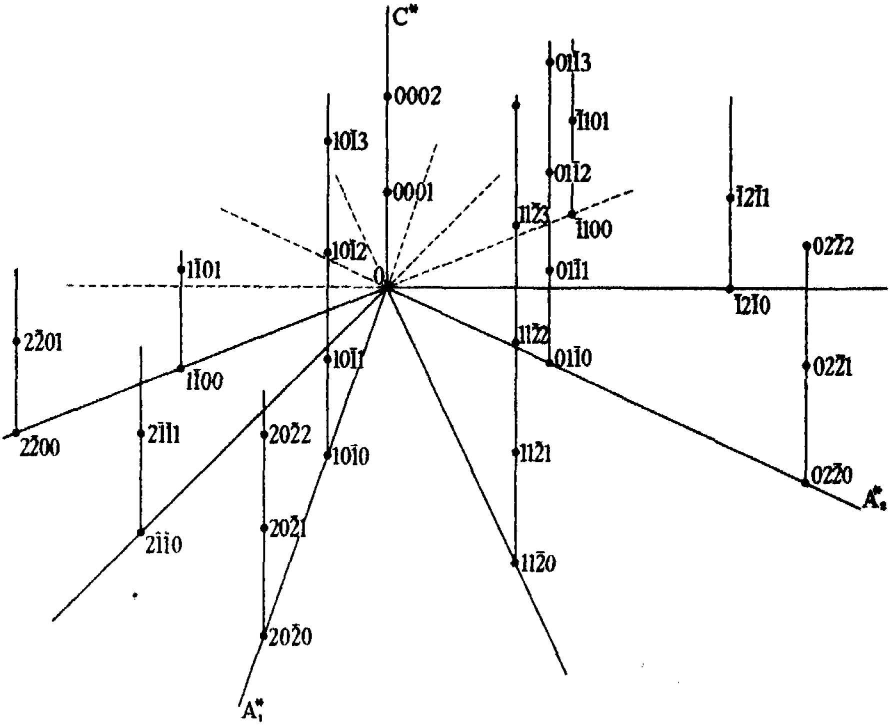


FIG. 1—Perspective drawing of reciprocal lattice of metallic cadmium.

Fig. 3 reproduces the pattern shown in fig. 2, Plate 3. The geometry of the pattern is easily understood from its hexagonal symmetry. The rings depicted in fig. 3 are due to CdO which has a face-centred cubic structure of lattice constant $a = 4{\cdot}70$ A. These rings serve as an accurate means of calibrating the pattern, and any radial movement of the spots relative to the rings, which is not less than 0·3 mm, can be detected.

Assuming that the oxide is identical in structure with bulk CdO, we have deduced the d_{hkl} values of the spots, *see* Table I. There is a close agreement between the observed d_{hkl} values and those calculated from the above structure for Cd. From the d_{hkl} values the indices of the spectra are obtained. It then appears that fig. 2 contains mostly spectra of type $h, k, \overline{h+k}, o$, and is identical with the hexagonal section of the reciprocal lattice containing the lines OA$_1$*, OA$_2$* in fig. 1. The various

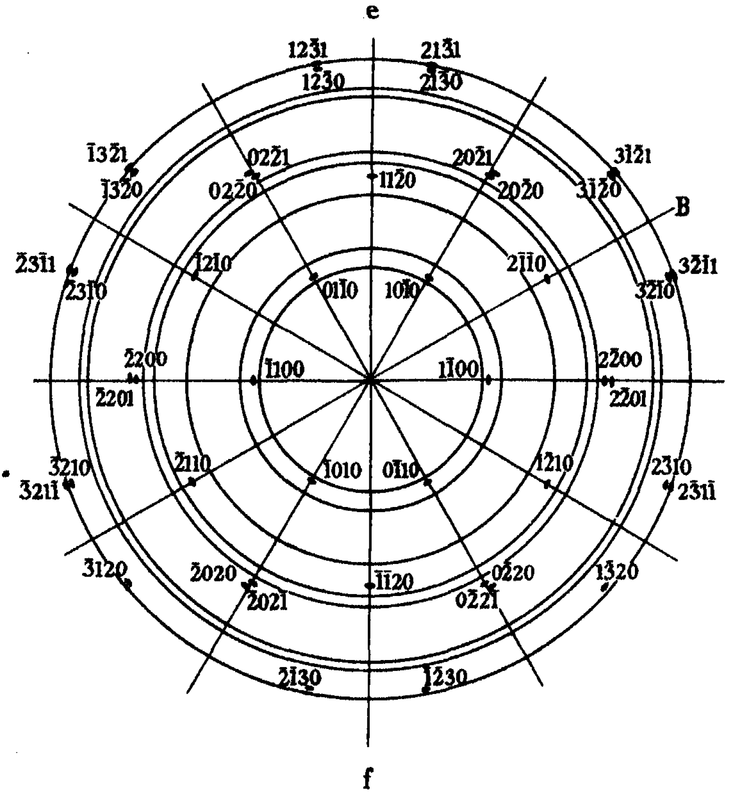


FIG. 3.

spectra observed are given in Table I, and in consulting this table, reference should be made to the reciprocal lattice, fig. 1.

In addition to the $h, k, \overline{h+k}, o$ spectra, certain other spectra of type $h, k, \overline{h+k}, 1$, appear which can only be seen faintly in the photograph.

The first question which concerns us is: could such a pattern be produced by a single perfect crystal? Secondly, can the pattern be explained as due to distortion of a single crystal (*i.e.*, to an aggregate of crystallites which as a whole resembles a single crystal)?

If the pattern has resulted from a single perfect crystal, the effect may be due to crystal size. But this appears to be impossible. For spectra of type $h, h, \overline{h+k}, o$ are visible over a circular area of diameter 7 cm, and this corresponds to a thickness $\ngtr$ 15 A if the crystal be a thin sheet cut normal to OC*. From the small value of this thickness it follows that the representative points in the reciprocal lattice must be extended along lines normal to the basal section. But if this were so, then we should expect the $\{20\bar{2}1\}$ spectra to be superimposed on the $\{20\bar{2}0\}$ spectra, and also the $\{21\bar{3}1\}$ on the $\{21\bar{3}0\}$ spectra. But these spectra appear quite distinct, and their d_{hkl} values show no measurable differences from the calculated values. Alternatively expressed, this superposition means that the rows of atoms normal to the sheet act as single diffracting units, so that spectra of type $l = 0$ and $l = 1$ have no separate existence.

TABLE I

Spectra observed	d_{hkl} observed	d_{hkl} calculated
$1\bar{1}00, \bar{1}100$; $10\bar{1}0, \bar{1}010$; $01\bar{1}0, 0\bar{1}10$	2·575	2·560
$1\bar{2}10, \bar{1}2\bar{1}0$; $2\bar{1}\bar{1}0, \bar{2}110$; $11\bar{2}0, \bar{1}\bar{1}20$	1·484	1·478
$2\bar{2}00, \bar{2}200$; $20\bar{2}0, \bar{2}020$; $02\bar{2}0, 0\bar{2}20$	1·283	1·280
$2\bar{2}0\overset{\pm}{1}, \bar{2}20\overset{\pm}{1}$; $20\bar{2}1, \bar{2}021$; $02\bar{2}1, 0\bar{2}21$	1·252	1·247
$2\bar{3}10, \bar{2}3\bar{1}0$; $3\bar{2}\bar{1}0, \bar{3}210$; $3\bar{1}\bar{2}0, \bar{3}120$ $21\bar{3}0, \bar{2}\bar{1}30$; $12\bar{3}0, \bar{1}\bar{2}30$; $\bar{1}3\bar{2}0, 1\bar{3}20$	0·971	0·966
$2\bar{3}11, \bar{2}3\bar{1}\bar{1}$; $3\bar{2}\bar{1}1, \bar{3}21\bar{1}$; $3\bar{1}\bar{2}1$, — ; $21\bar{3}1$, — ; $12\bar{3}1$, — ; $\bar{1}3\bar{2}1$, ·	0·958	0·952

The possibility that dynamical interaction alone is responsible for this pattern can be rejected immediately. For the $\{20\bar{2}1\}$ and $\{21\bar{3}1\}$ spectra would then have to be due to the interaction of beams of type $l = 0$, and this is impossible for electrons of 30 kV velocity, according to Bethe. In this case, also, one would not expect the d_{hkl} values to remain constant since the representative points of type $l = 1$ are not contained by the basal section of the reciprocal lattice. We conclude, therefore, that a distorted crystal or aggregate is chiefly responsible for the pattern.

The aggregate probably consisted of a number of flakes of Cd whose relative positions were such that they could be derived by rocking a single flake about all possible axes in its own plane. The plane of each flake is normal to the c axis. The radius vectors from 0 to the representative points $\{20\bar{2}1\}$ in the reciprocal lattice make equal angles of 12° 50′ with

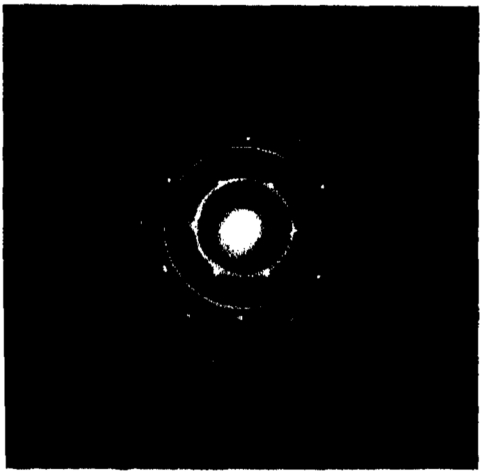


FIG. 2.

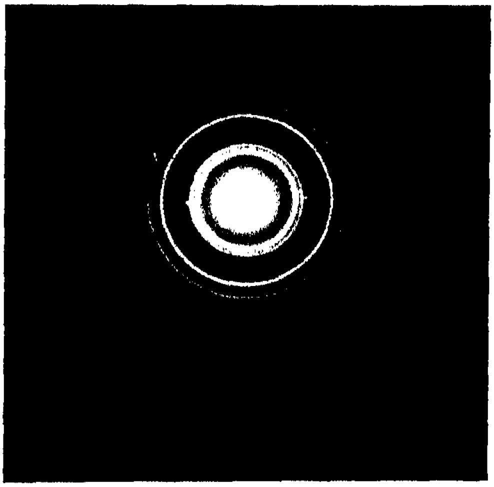


FIG. 5.

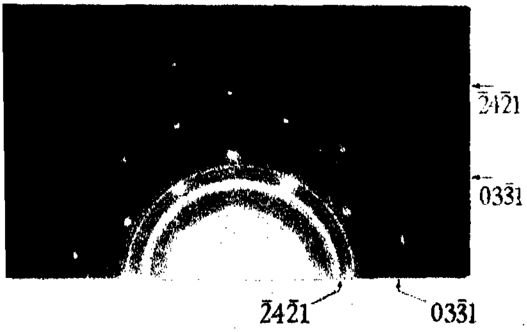


FIG. 4.

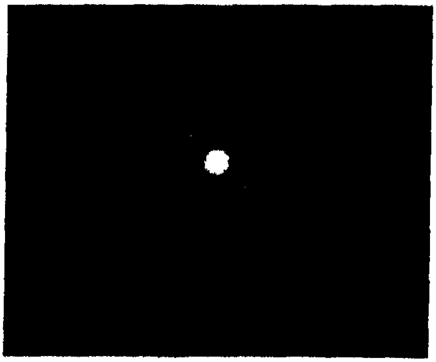


FIG. 7.

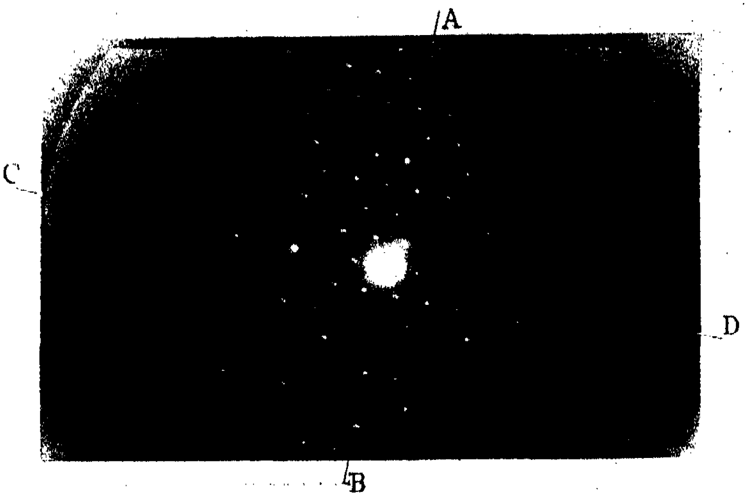


FIG. 10*a*.

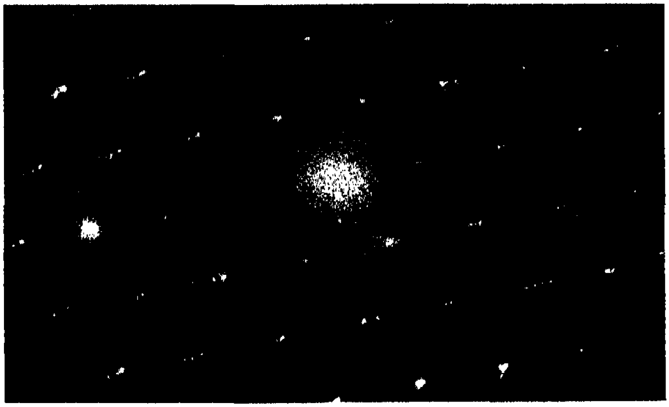


FIG. 10*b*.

the basal section: the corresponding angle for planes of type $\{21\bar{3}1\}$ is 9° 50′. The angle of warping of the flakes composing the aggregate was therefore not less than 12° 50′. It might be expected that such an aggregate would result in the spots being spread along possible Debye-Scherrer circles, and indeed such spreading may be observed in fig. 2, Plate 3. Calculation shows that the observed spreading is of the right order. It is necessary to point out here that an aggregate such as this may still give rise to sharp spots if the warping takes place most about some particular direction in the plane of the flake. It was found that the latter type of aggregate occurs by rotating a specimen about various axes in the electron-beam. The angular range over which the spots persisted was measured; all spots of the same type did not persist equally. During these rotation experiments there appeared no change in the radial positions of the spectra; the angles of persistence were roughly proportional to the structure factors, account being taken of the nature of the aggregate.

The pattern shown in fig. 4, Plate 3, was obtained by rotating the above specimen through an angle ~ 6° about an axis parallel to the line joining the $1\bar{1}00$, $\bar{1}100$ spectra in the reciprocal lattice. The intensities of the $\bar{2}020$ and $\bar{2}021$ spectra were thereby made approximately equal. The interesting result of this operation is the appearance of certain forbidden spectra, *i.e.*, spectra for which the structure factors F_{hkl} are zero, viz., the $\bar{3}031$, $03\bar{3}1$, $\bar{2}4\bar{2}1$, $\bar{4}221$ spectra. The most plausible explanation of these forbidden spectra is that they are due to successive reflections by different crystallites in the aggregate, *e.g.*, the $0\bar{3}31$ spectrum is due to successive reflections by the $01\bar{1}0$, $022\bar{1}$ planes, etc.

It may be shown that, if the $03\bar{3}1$ spectrum is due to successive reflections by the $01\bar{1}0$, $02\bar{2}1$ planes in different crystallites, then the $03\bar{3}0$, $03\bar{3}1$ spectra should be separated by 0·52 mm in fig. 4. On the other hand, if some crystals of the aggregate are so placed that the sphere of reflection coincides with the plane containing the origin and the representative points $\bar{1}100$, $03\bar{3}1$, etc., then interaction among the beams corresponding to representative points contained in this plane may give rise to the forbidden spectra observed. In the latter case we should not expect the positions of the forbidden spectra to differ appreciably from the positions predicted by the Bragg law; in particular, the separation of the $03\bar{3}0$, $0\bar{3}31$ spectra in this case should be 0·37 mm. The size of the $03\bar{3}1$ spot in fig. 4, Plate 3, prevents one from drawing a conclusion. It was not possible, therefore, to determine to what extent the energy of the forbidden ray is due to dynamical interaction.

Zinc

Crystallographic Data—Zn also has a structure resembling the hexagonal close-packed arrangement of atoms, and the dimensions of the hexagonal lattice are $a = 2{\cdot}65$ A, $c/a = 1{\cdot}864$. The drawing of the reciprocal lattice in fig. 1 applies equally well to Zn; the expression for F_{hkl} is the same as for Cd.

Fig. 6 shows in detail the spectra present in fig. 5, Plate 3, and the indices of the spectra are listed in Table II. This pattern lends itself to more precise treatment than the preceding patterns due to Cd. There is in fig. 5 approximate symmetry about the line AB; also the spectra $\bar{1}3\bar{2}1$, $\bar{2}3\bar{1}1$ appear strongly on one-half of the pattern but not on the other half.

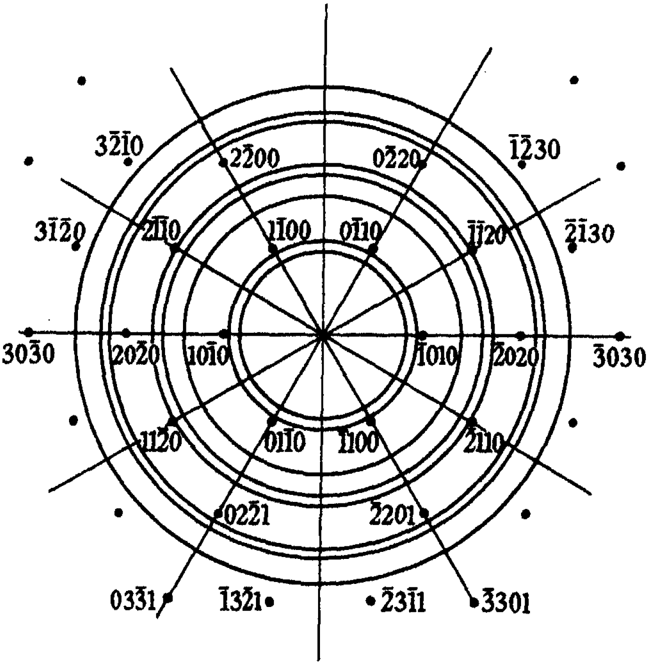


FIG. 6.

Hence it appears that the sphere of reflection must have passed approximately through the representative points $\bar{1}3\bar{2}1$, $\bar{2}3\bar{1}1$, and the origin. Assuming this, we have calculated the radial distances δ_{hkl} of the representative points from the sphere of reflection. The values of δ_{hkl} are given in Table II. From these data we see that when δ_{hkl} is $> 7 \times 10^6$ cm, no spectrum *hkl* appears. It now appears by putting $c^*/N_c = 7 \times 10^6$

cm, *see* p. 3, that the thickness of crystal required for the Bragg-Kirchner mechanism to be alone responsible for this pattern is $N_e < 3$ unit cells. Even taking into account dynamical interaction, this thickness is scarcely plausible. Further, it is reasonable to suppose that the effects due to dynamical interaction are of the same order for Cd as for Zn, hence if the pattern in fig. 2, Plate 3, due to Cd, is to be explained by means of a single crystal, then the thickness of crystal required will have to be very much less than 3 unit cells. This indicates that we are correct in attributing fig. 2 to distortion.

TABLE II

Spectra	δ_{hkl}	Intensities observed in fig. 5	F^2_{hkl}
$10\bar{1}0$	1×10^6	vvst	19×10^4
$01\bar{1}0$	5×10^6	w	19×10^4
$1\bar{1}00$	-4×10^6	wm	19×10^4
$\bar{1}2\bar{1}0$	12×10^6	o	23×10^4
$11\bar{2}0$	7×10^6	vw	23×10^4
$2\bar{1}\bar{1}0$	-3×10^6	vst	23×10^4
$1\bar{2}10$	-7×10^6	vw	23×10^4
$20\bar{2}0$	3×10^6	st	4×10^4
$02\bar{2}0$	12×10^6	o	4×10^4
$2\bar{2}00$	-6×10^6	vw	4×10^4
$02\bar{2}1$	-7×10^6	vw	11×10^4
$\bar{2}3\bar{1}1$	0×10^6	vst	6×10^4
$\bar{3}211$	-5×10^6	vw	6×10^4
$\bar{3}120$	10×10^6	o	2×10^4
$\bar{2}\bar{1}30$	1×10^6	vst	2×10^4
$\bar{1}\bar{2}30$	-4×10^6	wm	2×10^4
$1\bar{3}20$	-8×10^6	o	2×10^4
$\bar{3}030$	7×10^6	vw	5×10^4

Since, therefore, the thickness of the Zn crystal required is so small, and since, moreover, we observe patterns due to Zn showing all gradations between fig. 5, Plate 3, and the type obviously due to an aggregate, then it seems more reasonable to suppose that here also fig. 5 is due to an aggregate. This aggregate, however, is very nearly a single perfect crystal.

Lastly, we note the occurrence in fig. 5 of the forbidden spectra $\bar{3}301$, $03\bar{3}1$. Again we suggest that these are due to successive reflections by the $\bar{1}010$, $\bar{2}3\bar{1}1$, $10\bar{1}0$, $\bar{1}3\bar{2}1$ planes respectively, in different crystallites. They may, however, be due to dynamical interaction.

Another pattern frequently obtained from zinc and cadmium was that corresponding to the section of the reciprocal lattice containing the representative points $1\bar{1}00$, $10\bar{1}1$, etc., and the origin. In this pattern we observed the forbidden spectra $\{11\bar{2}1\}$; these could not be confused with any other spectra, for distortion was such that the $\{11\bar{2}0\}$, $\{11\bar{2}1\}$, $\{11\bar{2}2\}$ spectra occurred simultaneously. The following forbidden spectra have been found to occur in the Cd and Zn patterns: $\{30\bar{3}1\}$, $\{11\bar{2}1\}$, $\{22\bar{4}1\}$, $\{11\bar{2}3\}$.

Bismuth

Crystallographic Data—The structure of bismuth is face-centred rhombohedral of dimensions $a = 6 \cdot 540$ A, $\alpha = 87° \; 34'$ for the usual crystallographic axes. It is convenient, however, to use a well-known transformation by which the structure can be referred to hexagonal axes. The transformation formulae are

$$a_1 = a \sin \alpha/2 \qquad c_1 = 3a\sqrt{1 - \tfrac{4}{3} \sin^2 \frac{\alpha}{2}}$$

$$h_1 = \tfrac{1}{2}(k - l) \qquad k_1 = \tfrac{1}{2}(h - k) \qquad l_1 = h + k + l,$$

where the letters with the suffix refer to the hexagonal lattice. The structure factor $F_{h_1k_1l_1}$ is given by

$$F^2_{h_1k_1l_1} = 2f_{\mathrm{Bi}}^2 \left\{1 + 2\cos\pi(h + k + l)\cos\frac{\pi}{3}h - k - l)\right\}^2 \left[1 + \cos 2\pi\left\{\frac{2k + h}{3} + l\,(\tfrac{1}{6} - \cdot 026)\right\}\right].$$

Only those representative points in fig. 1 are now significant for which $F_{h_1k_1l_1} \not\equiv 0$.

Fig. 8 shows in more detail the pattern of spots appearing in fig. 7, Plate 3. Only those spectra occur for which the representative points lie on the plane passing the origin of the reciprocal lattice and the $10\bar{1}4$, $\bar{1}2\bar{1}0$ representative points. This corresponds to the electron beam having proceeded along an edge of the rhombohedral cell, which explains why the pattern is nearly a square array of points. The spectra occur over a circular area of diameter 5 cm, and using the formula $\phi_0 = 2\sqrt{\frac{2\lambda}{t}}$ this implies that if the Bragg-Kirchner mechanism alone were responsible for the pattern then the thickness of the diffracting crystal parallel to the edge of the rhomb was $\ngtr$ 27 A, which is not impossible. This, then, is precisely the type of pattern which one is tempted to assign to a single perfect crystal. But it was possible to show that an aggregate of crystals was again involved. This was made evident by rotating the specimen about a line parallel to the join of the $10\bar{1}4$, $\bar{1}01\bar{4}$ spectra, and taking photographs at intervals of 3° as the specimen was rotated. It was found in this way that the $\bar{1}2\bar{1}0$, $1\bar{2}10$ spectra persisted over 9° approximately on each side of the mean position. Similarly, by rotating about a line parallel to the join of the $\bar{1}2\bar{1}0$, $1\bar{2}10$ spectra, it was found that the $10\bar{1}4$, $\bar{1}01\bar{4}$ spectra persisted over $\nless$ 15° on each side of the mean position. Now since the value of F_{1120} is $>$ F_{1014} we should expect the $\{1120\}$ spectra to persist over a greater angle than the $\{1014\}$ spectra. This result cannot be explained by the Bragg-Kirchner mechanism nor by dynamical interaction.

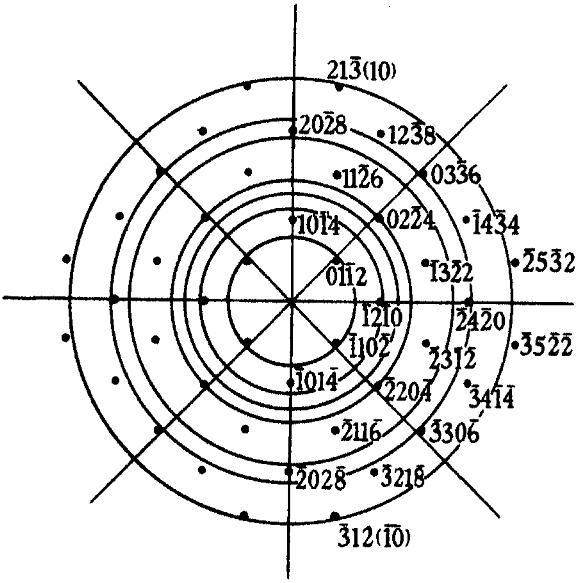


FIG. 8.

We must conclude that the representative points were extended in the reciprocal lattice because of imperfections in the diffracting crystal. It is more reasonable to suppose, therefore, that the pattern in fig. 7, Plate 3, is due chiefly to distortion. There was no evidence of radial movement of the spots during rotation.

Aluminium

The patterns obtained from aluminium were always obviously due to aggregates. The spots became arcs of Debye-Scherrer circles and these arcs subtended angles of 20° at the centre in some cases. Specimens of aluminium were also prepared by withdrawing a loop of copper wire from the molten metal *in vacuo*, in the diffraction chamber, but the crystals thus obtained were, if anything, less perfect.

Mica

Preparation and Experimental Procedure—The mica films were prepared by cleavage of a large crystal of muscovite mica. The object was not to repeat Kikuchi's experiments, but to see whether effects could be obtained similar to the results described for the metals. For this reason no care was taken to mount the films flat, and we looked to see whether any effects due to natural curvature of the film were revealed in the diffraction patterns. Each film was mounted on a brass disc perforated with a hole 1 mm in diameter. The thinnest part of the film was made to project over the hole, and in some cases did not entirely cover the opening, so that bending was possible.

Crystallographic Data for Muscovite Mica—The space lattice of muscovite mica is monoclinic, and of dimensions,

$$a = 5\cdot18 \text{ A}$$
$$b = 9\cdot02 \text{ A}$$
$$c = 20\cdot04 \text{ A}$$
$$\beta = 95^\circ\ 30'.$$

The axes of the crystal lattice are drawn in fig. 9*a* and their relation to the axes oa^*, ob^*, oc^* of the reciprocal lattice is shown. The dimensions of the reciprocal lattice are

$$a^* = \frac{1}{a \sin \beta} = 0\cdot194 \times 10^8 \text{ cm}$$

$$b^* = \frac{1}{b} = 0\cdot111 \times 10^8 \text{ cm}$$

$$c^* = \frac{1}{c \sin \beta} = 0\cdot05 \times 10^8 \text{ cm}$$

$$\beta^* = 84^\circ\ 30'.$$

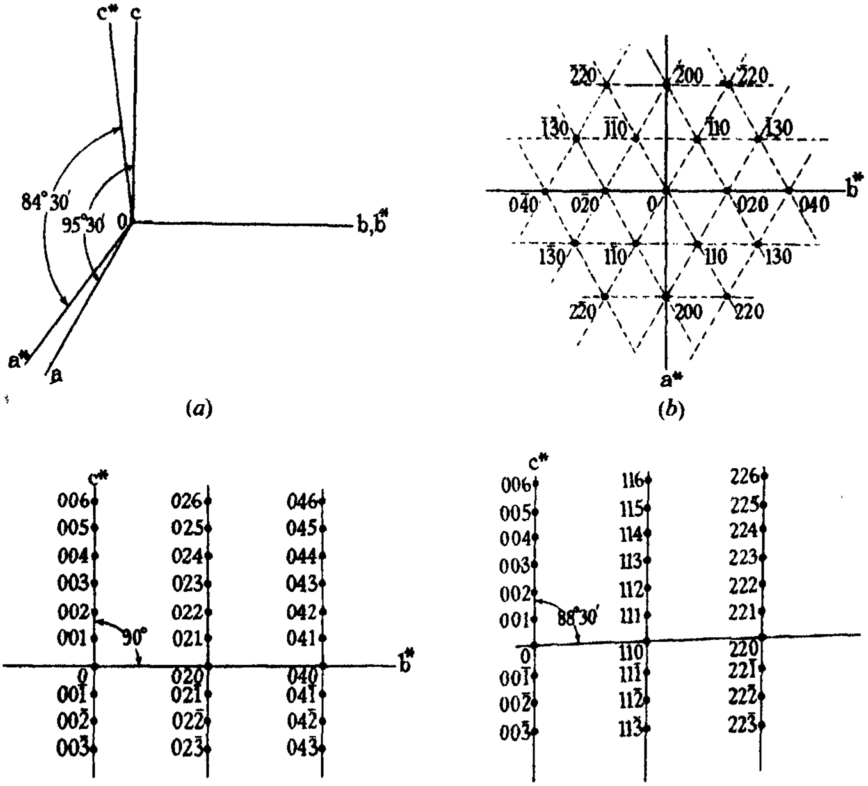
The molecules in the *ab* plane of the crystal lattice are arranged on a centred rectangular network, so that spectra for which $h + k$ is odd are forbidden. In fig. 9*b* a plan of the $a^*\ b^*$ plane of the reciprocal lattice is given, and since the ratio $a^*/b^* = 1\cdot749 = \sqrt{3}$ approximately the representative points in this plane form an equilateral network. The cleavage plane of mica is the *ab* plane in fig. 9*a*.

No measurements of the thicknesses of the mica films were made, but, from the work of Darbyshire,† we can estimate their thicknesses to be between

---

† 'Z. Kristallog.,' vol. 86, p. 313 (1933).

10^{-5} and 10^{-6} cm. Most of the patterns obtained resembled the Kikuchi patterns, *i.e.*, consisting of an equilateral network of spots, identical, so far as we can determine, with the a^*b^* plane, fig. 9*b*. Usually the area over which these patterns extended on the photographic plate could not be reconciled with the formula $\phi_0 = 2\sqrt{2\lambda/t}$.



(*a*) (*b*)

(*c*) (*d*)

Fig. 9.

Several photographs, however, indicated that besides the main pattern of equilaterally arranged spots, other spectra occurred which did not belong to this equilateral network, and in fig. 10*b*, Plate 4, an enlarged portion is reproduced of a photograph which is an extreme case of this. Similar photographs have also been obtained by Steinheil,† but no quantitative work has been carried out.

We ascribe the extra spots appearing in fig. 10 to the fact that the mica sheet was not perfectly flat, but bent about a single axis AB, fig. 10*a*, Plate 4, contained in the cleavage plane. This resulted in the extra spots appearing along lines parallel to CD. In order to describe the photograph we refer to fig. 9. Figs. 9*c* and 9*d* are sections of the reciprocal lattice containing the 001, 020 directions and the 001, 110 directions respectively.

---

† 'Ann. Physik,' vol. 5, p. 464 (1934).

The almost exact hexagonal symmetry of fig. 10 made it difficult to decide whether those spots lying on CD and nearest the central spot are the 020, $0\bar{2}0$ spectra or the 110, $\bar{1}\bar{1}0$ spectra, and so to decide whether warping had taken place about the O-200 direction as axis in the reciprocal lattice, or about the O-$\bar{1}30$ direction respectively. From fig. 9*c* we see that if warping had occurred about the O-200 direction then between the 020 and 040 spectra (which would lie along CD) there could have appeared a number of spots of type $02l$. In column 1, Table III, we give a list of these spots which would be expected to appear between 020 and 040 for an angle of warping of 60° approximately. In column 2 the d_{hkl} values calculated from the formula

$$d_{hkl} = \frac{b \sin \beta}{\sqrt{\frac{h^2b^2}{a^2} + \frac{l^2b^2}{c^2} - \frac{2hlb^2 \cos \beta}{ac} + k^2 \sin^2 \beta}}$$

are given, and in column 3 the observed values, calculated on the assumption that the innermost spot is the 020 reflection.

Referring now to fig. 9*d*, we see that if warping had occurred about the O-$\bar{1}30$ direction as axis, then between the 110 and 220 spots (which would now lie on CD) there could appear spectra of type $11l$. In Table III, column 4, a list of the $11l$ spectra together with calculated and observed d_{hkl} values are given. In this case the observed d_{hkl} values were evaluated on the assumption that the spacing of the innermost spot on CD was d_{110}.

The agreement between columns 5 and 6 is much better than between columns 2 and 3; hence it is evident that warping occurred about the O-$\bar{1}30$ direction as axis, and that the spots along CD are of the type $11l$, and not $02l$. By studying the actual distribution of the spots between the 220 and 330 spectra, the angle of warping was fixed to be 60° approximately. Since no spectra of the type $11l$ (*see* Table III) were observed, the warping was not symmetrical about the direction of the incident beam; the film must have been mounted with the cleavage plane mainly normal to the beam and the mica was bent in one direction only. This agreed with our direct observations of the specimen.

The spots forming the main pattern are quite distinct from one another, and the diffuse haloes visible in fig. 10, Plate 4, do not confuse the spectra on the original photographic plate. If the Bragg-Kirchner mechanism

Table III—Table of Results of Mica

Spectra 1	Calculated 2	Observed 3	Spectra 4	Calculated 5	Observed 6	Spectra 7	Calculated 8	Observed 9	Spectra 10	Calculated 11
020	4·510	4·510	110	4·476	4·476	220	2·238	2·244	110	4·476
		Assumed for calibration			Assumed for calibration					
021	4·398	4·315	111	4·282	4·283	221	2·204	2·223	111	4·447
022	4·109	3·978	112	3·962	3·949	222	2·146	2·146	112	4·216
023	3·732	3·590	113	3·577	3·563	223	2·069	2·069	113	3·865
024	3·344	3·188	114	3·201	3·164	224	1·981	1·970	114	3·477
025	2·987	2·874	115	2·862	2·852	225	1·886	1·875	115	3·109
026	2·675	2·569	116	2·568	2·551	226	1·788	1·785	116	2·782
027	2·408	2·379	117	2·317	2·360	227	1·692	1·699		
028	2·255	2·261	118	2·104		228	1·601	1·588		
			119	1·923		229	1·513			
			11(10)	1·768		22(10)	1·431			

had been of importance in producing this mica pattern we should have expected the spots of type 11*l* to have spread towards each other and so to have formed almost continuous lines across the pattern. But there was no evidence of this. The Bragg-Kirchner effect was therefore of secondary importance here, and we can say that, so far as the geometry of the pattern is concerned, the effects are just those to be expected if the electron beam behaves as an X-ray beam of equal wave-length.

In fig. 10 we also observe additional spots which do not belong to the main pattern. These spots lie on possible Debye-Scherrer circles for mica, and we ascribe them to small flakes of mica which have become detached from the main sheet. Although it was not possible to pick out any regular net patterns formed by these additional spots, yet definite information can be given about their occurrence symmetrically on either side of the central spot. It was found that the spots occurred in pairs symmetrically about the central spot over a circular area of 4 cm diameter approximately. This would indicate that these detached flakes were about 50 A, *i.e.*, two unit cells thick if these reflections were due to the effect of low resolving power.

The diffraction spots in fig. 10 are much sharper than the spot formed by the incident beam. It is possible that this is due very largely to the curvature of the film. The film acts as its own slit, for on account of the curvature the *hkl* planes are only in a position to reflect over a small portion of the whole film.

CONCLUSIONS

It appears, from the patterns examined here, that the spot patterns so often obtained in electron diffraction work are due chiefly to aggregates of crystallites and not to single crystals. The group of crystals composing the aggregate is such that, as a whole, the aggregate is approximately a single crystal. It has been found that the aggregate may deviate from the single crystal state considerably and yet that the spots of the pattern remain sharp with little tendency to spread along Debye-Scherrer circles. An explanation of this has been given. All gradations between patterns of sharp spots and patterns in which the spots are spread along the rings have been obtained. The results indicate further that in some cases the aggregate is a distorted, but continuous, single crystal rather than a group of disconnected fragments.

Forbidden spectra have been observed, and these we consider are most probably due to successive reflections by different crystallites in an aggregate. The fact that these forbidden spectra occur, shows that the scattering power of atoms for electrons is so great that not only will successive reflections by different crystallites occur but also that a diffracted beam set up in a single perfect crystal must be reflected within the same crystal. The energy of a beam diffracted according to the Bragg law must therefore be shared with that of the direct beam, *i.e.*, dynamical interaction to this extent is certainly large. This indicates that the present method of calculating the intensities of electron diffraction Debye-Scherrer circles must be modified; the first correction necessary because of dynamical interaction is the same as that dealt with by Darwin for X-ray diffraction.

Whilst other effects, due to dynamical interaction and low resolving power, may be present, we do not require these to explain any spot pattern that we have examined. We are seeking to obtain further information as to the importance of the latter from the Debye-Scherrer patterns due to the metals; for if the forbidden spots are due to successive reflections by different crystallites, then we shall not expect to observe forbidden rings but rather broad, circular bands. Dynamical interaction, on the other hand, could give rise to forbidden rings.

The writers wish to express their appreciation of the kind interest taken in this work by Professor W. L. Bragg, F.R.S., and Mr. R. W. James.

SUMMARY

The object of the research was to examine the spot patterns obtained when an electron beam of 30 kV velocity is transmitted through a crystalline film. The paper is divided into two sections. The first section is a brief review of the work relating to such patterns; the reciprocal lattice picture of electron diffraction is presented in a form suitable for discussing the results. In the second section a number of patterns representative of those obtained from cadmium, zinc, and bismuth are described. The crystals of these metals were obtained by withdrawing the surface oxide film from the molten metals in air and in nitrogen; these films produced Debye-Scherrer patterns due to the oxide and also spot patterns due to occluded metal crystals. The spot patterns were found to be due chiefly to distortion of the metal crystals. Forbidden spectra were observed. We also describe the results obtained by transmitting an electron beam through curved sheets of mica, and these again support the conclusion that the spot patterns are due chiefly to distortion of the diffracting crystal.

Experiments on Heavy Hydrogen

V—The Elementary Reactions of Light and Heavy Hydrogen. The Thermal Conversion of Ortho-Deuterium and the Interaction of Hydrogen and Deuterium

By ADALBERT FARKAS and LADISLAS FARKAS, Colloid Science Laboratory, Cambridge

(*Communicated by Eric K. Rideal, F.R.S.—Received April* 5, *Revised July* 10, 1935)

Since the discovery and preparation of deuterium, many of the chemical reactions of this heavy hydrogen isotope have been investigated and compared with the corresponding reactions of ordinary hydrogen.* It was soon recognized that in certain reactions† the differences observed between the rate of a reaction involving hydrogen molecules and of the corresponding reaction involving deuterium molecules (D_2) could be explained by the assumption that the different zero-point energies of the molecules H_2 and D_2 contribute a different amount of energy for the activation process. This was, in fact, predicted by Cremer and Polanyi, and by Eyring.‡ This difference in the zero-point energies of the initial states will cause a lower reactivity of heavy hydrogen as compared with light hydrogen. On the other hand, it was pointed out by Polanyi§ that the activated state also contains a certain amount of zero-point energy which is smaller for the deuterium complex than for the hydrogen complex and that therefore the energy of activation may appear to be lowered for reactions involving heavy hydrogen when compared with the corresponding reaction of light hydrogen.

The subject of the present paper is the investigation of the effect of replacing the hydrogen atoms and molecules by deuterium atoms and molecules on the rate of the interaction of hydrogen atoms and molecules according to the reaction

---

* Farkas, 'Naturwiss.,' vol. 22, pp. 614, 640, 658 (1934); Urey and Teal, 'Rev. Mod. Phys.,' vol. 7, p. 34 (1935); Farkas, "Ortho-hydrogen, Para-hydrogen, and Heavy Hydrogen," Cambridge University Press (1935).

† Farkas and Farkas, 'Naturwiss.,' vol. 22, p. 218 (1934); Rollefson, 'J. Chem. Phys.,' vol. 2, p. 144 (1934); Bach, Bonhoeffer, and Moelwyn-Hughes, 'Z. phys. Chem.,' B, vol. 27, p. 71 (1934).

‡ 'Z. phys. Chem.,' B, vol. 19, p. 443 (1932); 'Proc. Nat. Acad. Sci. Wash.,' vol. 19, p. 78 (1933).

§ 'Nature,' vol. 133, p. 26 (1934).

$$H + H_2 \rightleftharpoons H_2 + H. \qquad (1)$$

This particular reaction involves only one reactant and a single process, and is the simplest chemical reaction we know. It appeared, therefore, to be an especially suitable example for such a study, being also well understood experimentally, and, owing to its ideal simplicity, amenable to theoretical treatment.

In order to obtain some information of the effect of replacing the hydrogen atoms in this elementary reaction by deuterium, two reactions have been investigated: (1) the thermal re-conversion of ortho-deuterium to normal deuterium by the mechanism

$$D + D_2 \rightleftharpoons D_2 + D, \qquad (2)$$

and (2) the interaction of (light) hydrogen with deuterium according to $H_2 + D_2 = 2HD$ involving the steps

$$D + H_2 \rightleftharpoons HD + H, \qquad (3)$$

$$H + D_2 \rightleftharpoons HD + D. \qquad (4)$$

In comparing the reaction velocities we adopt the usual method of relating the reaction velocities to the collision numbers.

It has been pointed out recently* that this treatment is in general not adequate and is only applicable in special cases. The exact quantum-mechanical formula for the velocity constant of simple reactions contains mass, moment of inertia, vibrational frequencies, etc., of the reaction partners and of the activated complex. A direct application of this formula is, however, made difficult, since the data for the activated complex are not known sufficiently exactly, and it is mainly for this reason that we have chosen the method hitherto applied of comparing reaction velocities with collision numbers. It will be seen that the experimental results can be satisfactorily explained by this simplified treatment.†

For the sake of clarity, we propose to give a brief summary of the experimental results and of the theoretical considerations relating to the reaction $H + H_2 = H_2 + H$ before describing the actual experiments.

* Pelzer and Wigner, 'Z. phys. Chem.,' B, vol. 15, p. 445 (1932); Wigner, 'Z. phys. Chem.,' B, vol. 19, p. 203 (1933); Eyring, 'J. Chem. Phys.,' vol. 3, p. 107 (1935).

† [*Note added in proof, July* 10, 1935—The exact calculation of the velocity constants for the elementary reactions of light and heavy hydrogen will be given in a forthcoming paper by L. Farkas and Wigner. It may be stated at once that the velocity constants calculated according to the exact theory are in agreement with the present experimental data.]

1—THE REACTION $H + H_2 \rightleftharpoons H_2 + H$

It was shown by A. Farkas[*] that the thermal reconversion of parahydrogen (p-H_2) into normal hydrogen [containing 75% ortho-hydrogen (o-H_2) and 25% p-H_2] at 600–750° C proceeds according to the mechanism:

$$H + p\text{-}H_2 \rightleftharpoons o\text{-}H_2 + H, \tag{1A}$$

or denoting the orientation of the nuclear spin of the H-atoms by an arrow

$$\boxed{\uparrow + \uparrow}\downarrow = \uparrow\ \uparrow + \downarrow\,. \tag{1B}$$

The H-atoms result from the thermal dissociation of H_2 and their concentration corresponds to the thermodynamical equilibrium

$$H_2 \rightleftharpoons 2H. \tag{5}$$

The order of the reaction is 3/2, showing that H-atoms and H_2 molecules are actually involved.

The equilibrium concentration of ortho- and para-hydrogen (3 : 1) is established through the reaction from left to right in formula (1A) being three times quicker than that from right to left. The reaction rate can be represented by the equation

$$-\frac{dp\,[H_2]}{dt} = k'_1 p\,[H_2]\,[H] - k''_1\,.\,(100 - p)\,[H_2]\,[H], \tag{6}$$

which leads when integrated to

$$p - p_\infty = (p_0 - p_\infty)\, e^{-(k'_1 + k''_1)[H]t}, \tag{7}$$

where p, p_0 and $p_\infty = 25$ denote the p-H_2 percentages at the times t, $t = 0$ and $t = \infty$, [H] the concentration of H-atoms and $k'_1 = 3k''_1$. Introducing the excess concentration of para-hydrogen $u = p - p_\infty$ and $k_1 = k'_1 + k''_1$, we obtain

$$u = u_0 e^{-k_1[H]t}. \tag{7A}$$

The activation energy of the reaction (2) was found to be 5·5 kg cals,[†] whereas the apparent energy of activation calculated from the dependence of the reaction rate on temperature is 58 kg cals; but this latter figure contains, of course, also half the heat of dissociation of H_2, *i.e.*, 51 cals used for the production of the atoms.

---

[*] 'Z. Elektrochem.,' vol. 36, p. 782 (1930); 'Z. phys. Chem.,' B, vol. 10, p. 419 (1931).

[†] According to experiments of Geib and Harteck, and Farkas and Harteck,[‡] in the temperature region 20–300° C the energy of activation is somewhat larger.

[‡] Geib and Harteck, 'Z. phys. Chem.,' Bodensteinband 849 (1931); Farkas and Harteck, 'Z. phys. Chem.,' B, vol. 25, p. 257 (1934).

The collision efficiency [(Z) = number of molecules reacting/number of molecules colliding] for all interchanges, including also those which do not lead to a para-ortho-hydrogen interconversion,

$$e.g., \uparrow\downarrow + \uparrow = \uparrow + \downarrow\uparrow,$$

is the sum of the collision efficiencies for the ortho-para and for the para-ortho conversion and is given by

$$Z = Z_{p\rightarrow o} + Z_{o\rightarrow p} = Se^{-5500/RT}, \tag{8}$$

$S = 0\cdot07$ designating the steric factor of the interchange. The collision efficiency Z is related to the velocity constant $k_1 = k'_1 + k''_1$ (given in litre/mole sec) by the relation

$$Z = \frac{k}{6\cdot06 \times 10^{20}} \frac{1}{2\sqrt{2\pi}\left(\frac{d_1 + d_{11}}{2}\right)^2 \sqrt{\frac{M_1 + M_{11}}{M_1 . M_{11}} RT}}, \tag{9}$$

d_1, d_{11} being the diameters of H and H_2, M_1 and M_{11} the molecular weights of H and H_2 respectively. Using the numerical values for d_1, d_{11}, M_1, and M_{11}, and R we obtain†

$$Z = \frac{k_1}{1\cdot6 \times 10^{10}\sqrt{T}}, \tag{10}$$

k is the true velocity constant if we refer to all atomic interchanges, including also those which do not lead to ortho-para conversions. Table I summarizes the experimental results obtained in the investigation of the reaction $H + H_2 \rightleftharpoons H_2 + H$. The quantity given in the second column is the "time of half-change" defined as the time in which u_0 decreases to $u_0/2$ (*cf.* formula (7A)) and is related to the velocity constant k_1, by the formula

$$\tau = \frac{ln2}{[H]k_1}. \tag{11}$$

TABLE I

Temperature ° K	τ in sec at 10 mm Hg press.	(H) in mole/litre at 10 mm Hg press.	k_1 in litre/mole sec	Collision efficiency (Z)	Steric factor (S)
873	6600	$7\cdot54 \times 10^{-14}$	$1\cdot37 \times 10^{9}$	0·00287	0·068
930	850	$4\cdot74 \times 10^{-13}$	$1\cdot71 \times 10^{9}$	0·00350	0·069*
973	204	$1\cdot69 \times 10^{-12}$	$2\cdot00 \times 10^{9}$	0·00400	0·070
1023	46	$6\cdot31 \times 10^{-12}$	$2\cdot38 \times 10^{9}$	0·00463	0·068

* New measurements.

---

† Farkas, 'Z. Elektrochem.,' vol. 36, p. 782 (1930); 'Z. phys. Chem.,' B, vol. 10, p. 419 (1931).

These results lead us to make the picture of the interchange

$$H + H_2 \rightleftharpoons H_2 + H, \tag{1}$$

in which the activated state is represented by a symmetrical linear configuration of three hydrogen atoms, and the energy of activation is the energy required to form this triatomic complex by overcoming the repulsion of the hydrogen molecule exerted on the approaching atom.

It was shown by Eyring and Polanyi* that this exchange reaction can be visualized as the movement of a point on the energy mountain representing the total energy of three hydrogen molecules in a linear configuration as a function of their mutual distances. Fig. 1 shows this energy mountain, the (arbitrary) heights being represented by contour lines. The initial state is in the valley at the left-hand top corner, the distance between the atoms A and B being large and the atoms B and C being bound to a molecule. The reaction path leads from the initial valley to the final valley, passing the minimum height, and is indicated by a line of arrows.

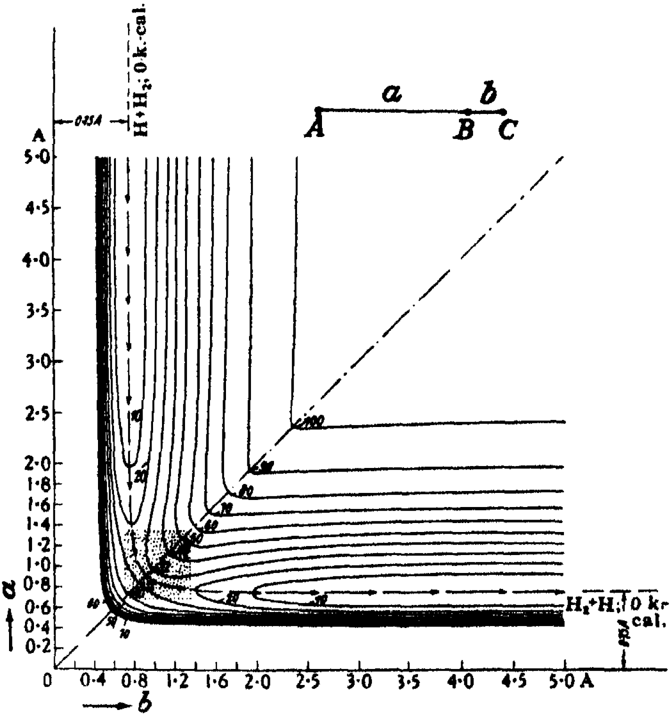


FIG. 1.

---

* 'Z. phys. Chem.,' B, vol. 12, p. 279 (1931).

A theoretical explanation for the steric factor occurring in the exchange reaction $H + H_2 \rightleftharpoons H_2 + H$ was given by Wigner.* It was shown by London† that the energy of activation for reactions of the type

$$A + BC \rightleftharpoons AB + C \tag{12}$$

is smallest if the atom approaches along the line joining the nuclei of the molecule. The dependence of the energy of activation (W) on the angle (δ) between the direction of motion and molecular axis is given according to Wigner by:

$$W \simeq W_0 + \frac{\delta^2 D}{10}, \tag{13}$$

(D = energy of dissociation of H_2, including the zero-point energy). Since the average energy of the molecule having more energy than W_0 is $W_0 + RT$, the maximum angle up to which the supplementary energy of activation can be supplied is given by

$$\frac{\delta^2 D}{10} = RT. \tag{14}$$

With $D = 108$ kg cals and $T = 1000°$ K, $\delta^2 = 0 \cdot 18$ or $\delta = 25°$. The steric factor for the reaction (1) is the ratio of the surface of the spherical cap subtended by this angle to the surface of the hemisphere, and thus we obtain $S = 1 - \cos 25° = 0 \cdot 09$ in satisfactory agreement with the experimental figure of 0·07. According to this theory the steric factor is slightly dependent on temperature, but this dependence might be neglected in the temperature region in which the experiments have been carried out.

The energy of activation for the reaction was theoretically estimated by London† and Eyring and Polanyi (*loc. cit.*). The obtained figures of 13 and 20 kg cals are too high compared with the experimental value of 5·5 kg cals.

2—THE REACTION $D + D_2 \rightleftharpoons D_2 + D$

The experimental arrangement for the investigation of the ortho-para-deuterium conversion was very simple. A quartz bulb of about 68 cc volume served as reaction vessel and was heated in an electric furnace. The temperature of the furnace was measured by means of a platinum-platinum-rhodium thermocouple and was found to be uniform in the middle of the furnace to within ± 3°. After having filled the reaction vessel with ortho-deuterium, from time to time small samples of about 0·003 cc (N.T.P.) were withdrawn from the vessel by means of a capillary lock, and their ortho-deuterium content analysed according to the micro-thermal conductivity method.*

---

* ' Z. phys. Chem.,' B, vol. 19, p. 203 (1932).

† Sommerfeld-Festschrift, p. 104, Hirzel (1928).

The deuterium was prepared by electrolysing 95% (or better) D_2O, separated from the oxygen by freezing the latter out at − 210° C, and finally purified by diffusion through a palladium tube.† During the process of preparation, the deuterium content of the gas diffusing through the palladium tube was measured and the gas was only collected when the deuterium content exceeded 98%. The normal deuterium obtained in this way consists of 2/3 ortho-deuterium and 1/3 para-deuterium.‡ It was converted into ortho-deuterium by adsorption and desorption on charcoal at the temperature of liquid hydrogen,† the equilibrium concentration at 20° K being 98% ortho-deuterium + 2% para-deuterium.§ Ortho-deuterium when kept in clean vessels is very stable and there was no reconversion to normal deuterium noticeable at room temperature even after 2–3 months.

The changed ortho-para-deuterium ratio in the sample treated with charcoal becomes apparent when the change of the thermal conductivity is measured. As usual, this was done by determining the resistance of an electrically heated, fine, platinum wire stretched in a small vessel filled with the gas to be analysed.‖ The values of the resistance for different hydrogen samples are given in Table II.

TABLE II

	ohms
Normal hydrogen	102·72
Para-hydrogen	105·30
Normal deuterium	111·47
Ortho-deuterium	113·04

On the basis of previous experiments¶ it was assumed that the observed changes in the resistances are proportional to the excess concentration of ortho-deuterium above the normal equilibrium ratio.

---

* Farkas, ' Z. phys. Chem.,' B, vol. 22, p. 344 (1933); Farkas and Farkas, ' Proc. Roy. Soc.,' A, vol. 144, p. 467 (1934).

† Farkas, Farkas, and Harteck, ' Proc. Roy. Soc.,' A, vol. 144, p. 481 (1934).

‡ Owing to the different spin and statistics the concentration of the ortho- and para-modifications and their dependence on temperature is different for D_2 when compared with H_2 (*cf.* also Farkas, Farkas, and Harteck, *loc. cit.*).

§ Johnston and Long, ' J. Chem. Phys.,' vol. 2, p. 389 (1934).

‖ Details of the method, *cf.* Farkas, Farkas, and Harteck (*loc. cit.*).

¶ Farkas, ' Z. phys. Chem.,' B, vol. 22, p. 344 (1933); Farkas, Farkas, and Harteck, ' Proc. Roy. Soc.,' A, vol, 144, p. 481 (1934).

During the experiments on the thermal conversion of ortho-deuterium, particular attention was given to the following points. In order to detect a possible catalytic activity of the walls of the reaction-vessel the known rate of reconversion of para-hydrogen was checked from time to time. It will be seen in fig. 3 that the rates for the para-hydrogen conversion obtained in this way are, within the limits of experimental error, in agreement with the original experiments of Farkas,* showing that the heterogeneous reaction occurring in the vessel used was negligible. In the course of these experiments it was found that there was a slow interchange between the hydrogen in the walls of the quartz vessel and the deuterium introduced into the vessel, and that thus the deuterium became gradually "lighter." Although this reaction proceeds very much more slowly† than the homogeneous conversion, it may affect the accuracy of the experiments on the latter process, since 1% of hydrogen in deuterium causes about the same change in the thermal conductivity of the sample as the reconversion of 10% ortho-deuterium into normal deuterium. Thus before the actual experiments were begun the vessel was thoroughly out-gassed and then left in contact for a considerable time with deuterium in order to replace in the walls all hydrogen by deuterium. This was necessary since it was found that the exchange was going on even at a stage when practically no hydrogen was given up. After such treatment no exchange could be observed during a period of an hour or two.

The progress of the thermal reconversion of ortho-deuterium and its dependence on temperature and pressure was found to be very similar to the conversion of para-hydrogen.

The rate of conversion is again given by

$$-\frac{do\,[D_2]}{dt} = k'_2 o\,[D_2]\,[D] - k''_2(100 - o)\,[D_2]\,[D]. \qquad (15)$$

Since in equilibrium $o_\infty = 66 \cdot 6$ and therefore $2k'_2 = k''_2$ integrating equation (15) we obtain

$$u_t = u_0 e^{-k_2 [D] t}, \qquad (16)$$

where u denotes the excess concentration of ortho-deuterium $u = o - o_\infty$ and $k_2 = k'_2 + k''_2$. k_2 is the proper velocity constant for the exchange reaction

$$D + D_2 \rightleftharpoons D_2 + D, \qquad (2)$$

---

* 'Z. Elektrochem.,' vol. 36, p. 782 (1930); 'Z. phys. Chem.,' B, vol. 10, p. 419 (1931).

† Farkas and Farkas, 'Trans. Faraday Soc.,' vol. 31, p. 821 (1935).

if we consider all interchanges, including also those not involving an ortho-para conversion.

Fig. 2 shows the progress of the reaction in three different runs, log u/u_0 being plotted against the time. The times of half-change (τ_P) observed in different runs in the temperature range 630–708° C and at pressures (P) 8 to 74 mm Hg are shown in Table III, together with the times of half-change at a pressure of 10 mm calculated according to the well-known relation

$$\frac{\tau_{10}}{\tau_P} = \sqrt{\frac{P}{10}}. \tag{17}$$

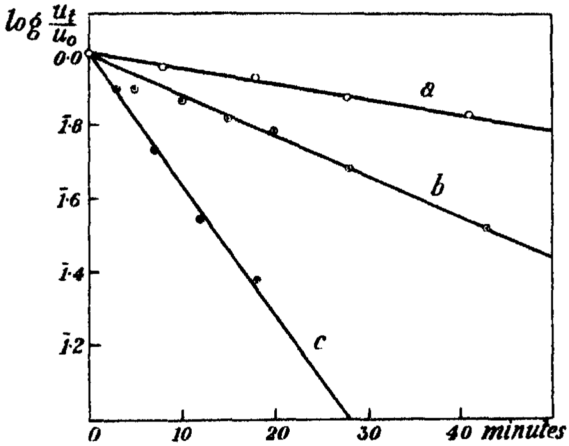


FIG. 2—*a*, 610° C, 32 mm Hg pressure; *b*, 654° C, 29 mm Hg pressure; *c*, 698° C, 18 mm Hg pressure.

TABLE III

Temperature °K	1000/*T*	Pressure (*p*) in mm Hg	Time of half-change (τ_P) in sec	Time of half-change at 10 mm Hg (τ_{10}) in sec	Log (τ_{10})
981	1·019	12	444	487	2·688
973	1·028	8	570	510	2·708
971	1·030	18	515	692	2·840
969	1·032	13	630	718	2·856
945	1·058	14	1115	1320	3·120
933	1·072	9	1680	1595	3·202
933	1·072	11	1500	1580	3·198
932	1·073	31	1020	1790	3·254
931	1·074	15	1740	2120	3·326
931	1·074	74	960	2610	3·417
927	1·079	29	1560	2650	3·424
903	1·107	29	3300	5610	3·749

Fig. 3 shows the dependence of the time of half-change, at 10 mm pressure, on temperature (curve *b*) compared with that for the para-hydrogen conversion (curve *a*), the logarithm of τ_{10} being plotted against the reciprocal of the absolute temperature.

From these data it follows that the thermal reconversion of ortho-deuterium proceeds in the temperature region 600–750° C 2·4–2·5 times more slowly than the para-hydrogen conversion at the same temperature and pressure. Since the measured velocity constants $k_1 = k'_1 + k''_1$ and $k_2 = k'_2 + k''_2$ for the ortho-para conversion of hydrogen and deuterium respectively are the true reaction velocity constants for the reactions

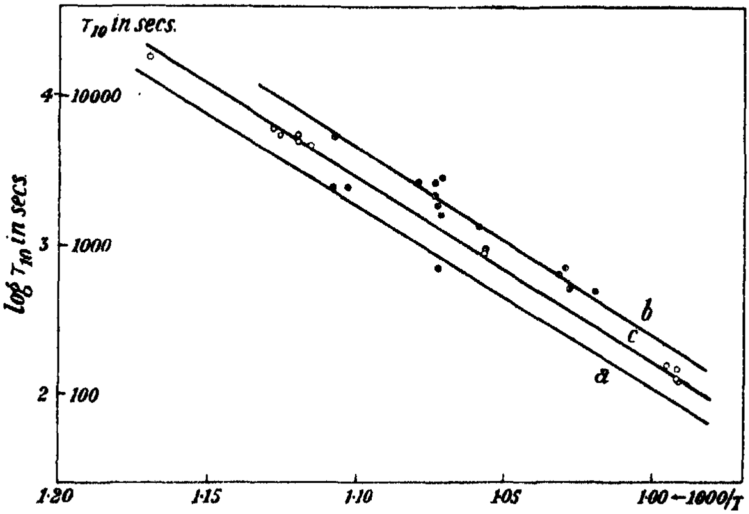


FIG. 3—*a*, Conversion of para-hydrogen; *b*, conversion of ortho-deuterium; *c*, interaction of hydrogen and deuterium.

$$H + H_2 \rightleftharpoons H_2 + H \qquad (1)$$

and

$$D + D_2 \rightleftharpoons D_2 + D, \qquad (2)$$

the ratio 2·4 is also characteristic of the relative rates of the reactions (1) and (2).

This difference in the rates of reaction is due to several causes. According to formulae (8), (9), and (11), the ratio of the times of half-change for the reactions (1) and (2) is given by

$$\frac{\tau_1}{\tau_2} = \frac{[D]\,k_2}{[H]\,k_1} = \frac{[D]\,S_2 e^{-W_2/RT}\left(\frac{d_2 + d_{22}}{2}\right)\sqrt{\frac{M_2 + M_{22}}{M_2 \,.\, M_{22}}}}{[H]\,S_1 e^{-W_1/RT}\left(\frac{d_1 + d_{11}}{2}\right)^2\sqrt{\frac{M_1 + M_{11}}{M_1 \,.\, M_{11}}}}, \qquad (18)$$

where the index 1 refers to the reaction (1) and the index 2 to the reaction (2).

From the theoretical explanation of Wigner for the steric factor one would expect that the reactions (1) and (2) would have the same steric factor, *i.e.*, $S_1 = S_2$.* If we assume, furthermore, that the collision diameters for light and heavy hydrogen molecules and light and heavy hydrogen atoms are the same we obtain the formula

$$\frac{\tau_1}{\tau_2} = \frac{[\mathrm{D}]\, e^{-\mathrm{W}_2/\mathrm{RT}}}{\sqrt{2}\,[\mathrm{H}]\, e^{-\mathrm{W}_1/\mathrm{RT}}}, \tag{19}$$

where the factor $\sqrt{2}$ accounts for the smaller collision frequency between D_2 and D compared with that between H_2 and H.

The D-atom concentration in thermal equilibrium with D_2 will be different from the H-atom concentration in H_2 at the same pressure and temperature through the different heats of dissociation and chemical constants of light and heavy hydrogen.† The ratio of the dissociation constants

$$\mathrm{K}_{\mathrm{H}_2} = \mathrm{P}_{\mathrm{H}}{}^2/\mathrm{P}_{\mathrm{H}_2}$$

and

$$\mathrm{K}_{\mathrm{D}_2} = \mathrm{P}_{\mathrm{D}}{}^2/\mathrm{P}_{\mathrm{D}_2}$$

is given up to 1000° K by

$$\log \frac{\mathrm{K}_{\mathrm{H}_2}}{\mathrm{K}_{\mathrm{D}_2}} = \frac{\Delta \mathrm{E}}{4\cdot 57\mathrm{T}} + 2i_{\mathrm{H}} - i_{\mathrm{H}_2} - 2i_{\mathrm{D}} + i_{\mathrm{D}_2}. \tag{20}$$

The difference in the heats of dissociations ΔE is equal to the difference in the zero-point energies of H_2 and D_2 1793 cals, and the chemical constants have the following values (nuclear spin contribution excluded):

$$i_{\mathrm{H}_2} = -\,3\cdot 36 \qquad i_{\mathrm{D}_2} = -\,2\cdot 61$$
$$i_{\mathrm{H}} = -\,1\cdot 28 \qquad i_{\mathrm{D}} = -\,0\cdot 83.$$

From the formula

$$\log \mathrm{K}_{\mathrm{H}_2}/\mathrm{K}_{\mathrm{D}_2} = 1793/4\cdot 57\mathrm{T} - 0\cdot 15$$

we obtain the ratios

$$[\mathrm{H}]/[\mathrm{D}] = \sqrt{\mathrm{K}_{\mathrm{H}_2}/\mathrm{K}_{\mathrm{D}_2}}.$$

T° K		[H]/[D]
900	....	1·39
950	....	1·36
1000	....	1·32

---

* *See*, however, L. Farkas and Wigner (*in preparation*).

† Johnston and Long, ' J. Chem. Phys.,' vol. 2, p. 389 (1934).

If we allow for this difference in the hydrogen atom concentration we find for 950° K

$$\frac{e^{-W_2/RT}}{e^{-W_1/RT}} = \frac{1 \cdot 41 \times 1 \cdot 36}{2 \cdot 5} = 0 \cdot 76,$$

or $W_2 - W_1 = 510$ cals and is the difference in the energies of activation involved in the reactions

$$D + D_2 \rightleftharpoons D_2 + D$$

and

$$H + H_2 \rightleftharpoons H_2 + H.$$

This difference in the activation energies is remarkably low compared with the difference in the zero-point energies of the molecules H_2 and D_2, which is 1793 cals.

If the energy level of the activated state in the reaction (2) were the same as that in the reaction (1) it would be expected that the energies of activation for the reactions (1) and (2) would differ by the energy difference in the initial state, *i.e.*, by the difference in the zero-point energies of the molecules H_2 and D_2 $(E_{H_2}^0 - E_{D_2}^0)$. The explanation for this not being so is the following. The activated complex consisting of three hydrogen (or deuterium) atoms in a linear configuration is to some extent a molecule (*see below*) and contains as such a certain amount of zero-point energy, due to a stable vibration. The zero-point energy for the complex DDD will be smaller than that for HHH on account of the smaller frequency associated with the heavier masses. Consequently the energy level for the activated state involved in the reaction

$$D + D_2 \rightleftharpoons D_2 + D$$

will be lower than that involved in the reaction

$$H + H_2 \rightleftharpoons H_2 + H.$$

Since the activation energy is the energy difference between the activated state and the initial state, it is evident that the existence of a zero-point energy in the activated state compensates to a certain degree the effect of the initial zero-point energy on the activation energy.

The energy diagram for the reactions (1) and (2) is shown in fig. 7. The absolute values of the zero-point energies in the activated state were calculated from the relations*

$$W_2 - W_1 = E_{H_2}^0 - E_{D_2}^0 - (E_{HHH}^0 - E_{DDD}^0) = 0 \cdot 51 \text{ kg cals.}$$

$$E_{H_2}^0 - E_{D_2}^0 = 1 \cdot 793 \text{ kg cals.}$$

$$\frac{E_{HHH}^0}{E_{DDD}^0} = \sqrt{2}.$$

The result being $E_{HHH}^0 = 4 \cdot 37$ and $E_{DDD}^0 = 3 \cdot 09$ kg cals.

---

* The ratio of the corresponding frequencies in the "molecules" HHH and DDD was assumed to be $\sqrt{2}$, *i.e.*, the same as that of the frequencies in the molecules H_2 and D_2. A more exact calculation is given by L. Farkas and Wigner (*in preparation*).

The zero-point energies for the activated complexes HHH and DDD, obtained from these experiments, are of the same order of magnitude as those one would expect from the calculation of the vibrations in the molecule HHH. It should be emphasized, however, that the complex HHH is not a molecule in the usual sense. The essential difference between a stable molecule and the activated complex is that the latter is not stable in a displacement corresponding to an antisymmetric vibration, but only in a displacement corresponding to a symmetric vibration (*cf.* fig. 4 showing the three normal vibrations of a linear triatomic molecule).

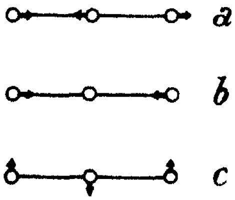


FIG. 4—*a*, Unsymmetric vibration; *b*, symmetric vibration; *c*, deformation vibration.

This is the case, since the motion corresponding to the antisymmetric vibration occurs in the direction of the reaction path shown in fig. 1, and the energy surface has a negative curvature in this direction. On the other hand, the motion corresponding to the symmetrical vibration is perpendicular to the reaction path.

Nevertheless, in calculating the vibrational frequency we shall consider the activated complex as an ordinary molecule, but allow for its instability by including only the symmetrical vibration.* This sort of calculation implies that the cross-section of the energy surface of the three hydrogen atoms along the symmetry axis, fig. 1, is similar to the corresponding cross-section through the energy surface for a stable triatomic molecule in which the bonds are of the same strength as those in the activated complex.

It is an empirical fact that in similar molecules the force constants (f) defining the vibrations are nearly proportional to the strength of the corresponding bond, *e.g.*, CO : force constant $18 \cdot 6 \times 10^5$ dyne/cm, heat of dissociation 248 kg cals; CO_2: force constant $13 \cdot 5 \times 10^5$ dyne/cm, heat of dissociation 188 kg cals. If we assume accordingly that the force constant for HHH with a heat of dissociation of 48 kg cals† is $2 \cdot 5 \times 10^5$

---

* [*Note added in proof, July* 10, 1935—The influence of the antisymmetric vibration on the zero-point energy of the activated complex is discussed by Farkas and Wigner.]

† The heat of dissociation of one H in HHH is half the whole energy in HHH, *i.e.*, energy in H_2 − energy of activation = 96 kg cals.

dyne/cm, since it is 5 × 10^5 dyne/cm for H_2, we obtain, according to the general relations valid for the triatomic molecules, the frequencies*

$$p_1^2 = \frac{4\pi^2 c^2 \omega_1^2}{N_L} = 5\cdot863 \times 10^{-2}\,\omega_1^2 = f.\frac{M_1 + M_2 + M_3}{M_2} \tag{21}$$

$$p_2^2 = 5\cdot863 \times 10^{-2}\,\omega_2^2 = f/M_2 \tag{22}$$

(c = velocity of light, N_L = Avogadro's number, M_1, M_3 = atomic weights of the lateral atoms, M_2 = atomic weight of the central atom in the triatomic molecule).

From these relations follows:

$$\omega_1 = 3580 \text{ cm}^{-1}$$

$$\omega_2 = 2060 \text{ cm}^{-1}.$$

If we neglect the third vibration, which is deflection vibration with a much smaller frequency, and bear in mind that the complex HHH is untenable to a displacement corresponding to ω_1, we find the zero-point energy of HHH to be $\frac{1}{2} \times 2060 \times 2\cdot842 = 2\cdot9$ kg cals† compared with the experimental figure of 4·37 kg cals.

3—The Interaction of Hydrogen and Deuterium

The experimental arrangement and method used for the investigation of the interaction

$$H_2 + D_2 \rightleftharpoons 2HD \tag{23}$$

was exactly the same as that used in the study of the ortho-deuterium reconversion. A mixture of about 50% H_2 + 50% D_2 was introduced in the reaction vessel and the amount of HD formed during the reaction was estimated by means of the micro-thermal conductivity‡ method, the resistance values being, *e.g.*,

		ohms
	H_2	107·82
52·5% D	$H_2 + D_2$	112·32
	$H_2 + D_2$ + HD in equilibrium	113·82
	D_2	116·37

---

* Kohlrausch, "Smekal-Raman Effekt," Springer (1931).

† Dr. B. Topley (private communication) has deduced a similar value for the zero-point energy of the activated complex from the form of the potential surface, *cf.* fig. 1, representing the energy of three hydrogen atoms as calculated by Eyring and Polanyi. 'Z. phys. Chem.,' B, vol. 12, p. 279 (1931).

‡ Farkas, 'Z. phys. Chem.,' B, vol. 22, p. 344 (1933); Farkas and Farkas, 'Proc. Roy. Soc.,' A, vol. 144, p. 467 (1934).

(The D_2 conductivity cell actually used was not the same as that used in the previous experiments, the former being more sensitive to changes in ortho-para-hydrogen concentrations, but the latter one better for changes in HD concentration.) The resistance value of the sample to be analysed was always compared with that of the original sample and with that of the same sample after equilibrium had been established on a hot nickel filament. Table IV shows the resistance values obtained in two experiments at 675° C and 4·5 mm Hg pressure and at 614° C and 28 mm Hg pressure.

TABLE IV

Temperature °C	Pressure in mm Hg	Time in minutes	Resistance value in ohms		ΔΩ	HD defect in %
			In sample	After equilibrium established		
673	4·5	0	112·32	113·82	1·50	100·0
		15	112·89	113·86	0·97	64·8
		30	113·29	113·88	0·59	38·4
		52	113·53	113·82	0·27	18·6
614	28	0	112·29	113·84	1·55	100·0
		24	112·77	113·90	1·19	76·5
		49	113·00	113·85	0·85	54·8
		74	113·15	113·83	0·68	43·8

The change in the resistance value after the equilibrium had been established was taken as a measure of the missing amount of HD, and in what follows the HD defect will be given as a percentage to characterize the progress of the reaction. Thus there is 100% defect at the beginning, 0% at the end of the reaction, and 50% at the time of half-change.

In order to avoid any separation of the hydrogen isotopes by diffusion,[*] the samples were taken with special care by suitable systems of locks, which allowed us to introduce the right amount of gas into the conductivity cell and to avoid unnecessary pumping through capillaries and taps.

The progress of the reaction $H_2 + D_2 = 2HD$ with time, and the dependence of its rate on pressure and temperature was again found to be very similar to that of the para-hydrogen conversion.

The progress of reaction can be represented by the equation

$$[HD]_\infty - [HD]_t = [HD]_\infty e^{-kt}, \tag{24}$$

where $[HD]_t$ and $[HD]_\infty$ denote the concentration of HD at the time t

[*] Farkas and Farkas, 'Proc. Roy. Soc.,' A, vol. 144, p. 467 (1934).

and at equilibrium, thus introducing the HD defect as $m_t = [\mathrm{HD}]_\infty - [\mathrm{HD}]_t$ we have

$$m_t = m_0 \, e^{-kt}. \tag{24A}$$

Fig. 5 shows three runs at 673° C, log m_t/m_0 when plotted against the time giving a straight line.

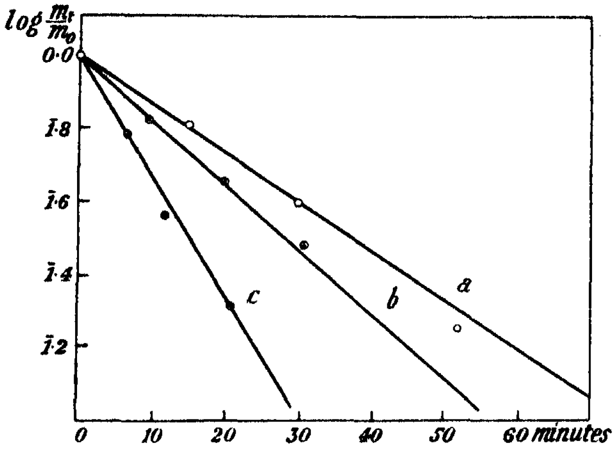


Fig. 5—Temperature 673° C, mm Hg pressure: *a*, 4·5 mm; *b*, 8 mm; *c*, 32 mm.

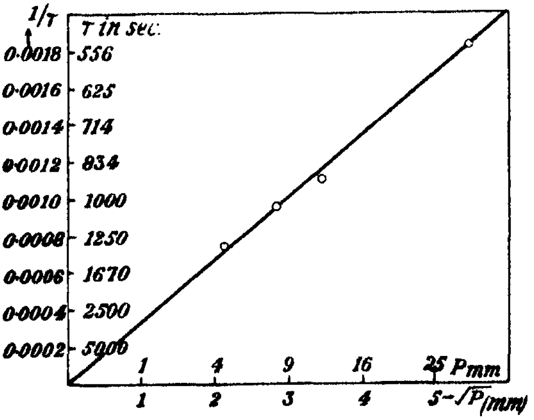


Fig. 6.

The order of the reaction was found to be 3/2, since the reciprocal times of half-change were proportional to the square root of the pressure (*see* fig. 6). This result shows that the interaction of hydrogen and deuterium actually takes place according to the atomic mechanism

$$\mathrm{D} + \mathrm{H}_2 \rightleftharpoons \mathrm{HD} + \mathrm{H}, \tag{3}$$

and

$$H + D_2 \rightleftharpoons HD + D, \tag{4}$$

whereas the molecular mechanism

$$H_2 + D_2 = 2HD \tag{25}$$

is negligible. This result is worth noting and for the following reason. This molecular reaction is prohibited for the ortho-para-hydrogen and deuterium interconversion, since the latter reactions involve a change of the symmetry with regard to the nuclear spins; but it might have appeared in the interaction of hydrogen and deuterium in which no change of symmetry takes place. But apparently the molecular mechanism has a higher energy of activation than 60 kg cals and occurs, if at all, only at higher temperatures. (In a previous note* it was reported that the order of reaction lies between 3/2 and 2. In view of the present results we consider this was due to a slight catalytic reaction occurring on the walls.)

TABLE V

Temperature °K	1000/T	Pressure (P) in mm Hg	Time of half-change (τ_P) in sec	Time of half-change at 10 mm Hg (τ_{10}) in sec	Log τ_{10}
1008	0·993	4·0	196	124	2·094
1008	0·993	4·5	219	147	2·167
1008	0·993	8	135	121	2·083
1004	0·996	3	288	158	2·198
946	1·057	4·5	1330	892	2·950
946	1·057	8	1038	928	2·968
946	1·057	12	900	985	2·993
946	1·057	32	546	976	2·990
897	1·116	15·5	3720	4630	3·666
893	1·120	10	4920	4920	3·692
893	1·120	10·5	5400	5540	3·744
889	1·126	17	4170	5440	3·735
887	1·128	28	3600	6020	3·780
855	1·170	29·5	10000	17200	4·235

The temperature dependence of the time of half-change at 10 mm pressure, calculated from the experimental results at different pressures according to the results shown in Table V, is given in fig. 3, curve *c*. It will be seen that in the temperature region 600–700° C the interaction $H_2 + D_2$ proceeds 1·5 times more slowly than the para-hydrogen conversion.

---

* Farkas and Farkas, 'Nature,' vol. 132, p. 894 (1933).

In order to calculate from these data the collision efficiencies and heats of activation for the different reactions involved in the interaction of hydrogen and deuterium, we must consider the kinetic equations

$$\underset{k_{d_1}}{H_2} \rightleftharpoons \underset{k_{r_1}}{2H} \qquad (26)$$

$$\underset{k_{d_2}}{D_2} \rightleftharpoons \underset{k_{r_2}}{2D} \qquad (27)$$

$$\underset{k_{d_3}}{HD} \rightleftharpoons \underset{k_{r_3}}{H + D} \qquad (27\text{A})$$

$$D + \underset{k_3}{H_2} \rightleftharpoons \underset{k_5}{HD} + H \qquad (3)$$

$$H + \underset{k_4}{D_2} \rightleftharpoons \underset{k_6}{HD} + D \qquad (4)$$

$$\frac{d\,[HD]}{dt} = k_3\,[H_2]\,[D] - k_5\,[HD]\,[H] + k_4\,[D_2]\,[H] - k_6\,[HD]\,[D] + k_{r_3}\,[H]\,[D] - k_{d_3}\,[HD] \qquad (28)$$

$$\frac{d\,[H]}{dt} = 2k_{d_1}\,[H_2] - 2k_{r_1}\,[H]^2 + k_3\,[H_2]\,[D] - k_5\,[HD]\,[H] + k_{d_3}\,[HD] - k_{r_3}\,[H]\,[D] + k_6\,[HD]\,[D] - k_4\,[D_2]\,[H] \qquad (29)$$

$$\frac{d\,[D]}{dt} = 2k_{d_2}\,[D_2] - 2k_{r_2}\,[D]^2 - k_3\,[H_2]\,[D] + k_5\,[HD]\,[H] + k_{d_3}\,[HD] - k_{r_3}\,[H]\,[D] - k_6\,[HD]\,[D] + k_4\,[D_2]\,[H]. \qquad (30)$$

k_{d_1}, k_{r_1}, k_{d_2}, k_{r_2}, k_{d_3}, and k_{r_3} are the velocity constants for the dissociation and recombination of the three hydrogens respectively, and k_3, k_4, k_5, and k_6 the velocity constants of the individual reactions between atoms and molecules.

[*Note added in proof, July* 10, 1935—The velocity constants for the forward and backward reactions are connected with each other, their ratio being the equilibrium constant of the corresponding reaction. Thus $k_{d_1}/k_{r_1} = K_{H_2}$ (*cf.* p. 134), $k_{d_2}/k_{d_2} = K_{D_2}$, etc. The equilibrium constants for the reactions (3) and (4) are at 1000°K $k_3/k_5 = 2{\cdot}61$ and $k_4/k_6 = 1{\cdot}49$ respectively.]

Thus the state of affairs in the interaction of H_2 and D_2 is much more complicated that in the two preceding reactions

$$H + H_2 \rightleftharpoons H_2 + H \quad \text{and} \quad D + D_2 \rightleftharpoons D_2 + D$$

for the following reasons. (1) The interaction of H_2 and D_2 involves two reactions. (2) The H- and D-atom concentration is disturbed by the exchange reactions (3) and (4) in which hydrogen atoms are exchanged for deuterium atoms and vice versa.

From the equations (29) and (30) we obtain the stationary [H] and [D] concentrations in the following way. We can assume that the sum [H] + [D] remains unchanged during the reaction since the [H] and [D] concentrations are altered mainly by the exchange reactions (3) and (4), whereas the rate of dissociation and recombination is negligible. Thus we may put

$$\frac{d[\mathrm{H}]}{dt} + \frac{d[\mathrm{D}]}{dt} = 0, \tag{31}$$

and obtain, assuming that $k_{r_1} = k_{r_2} = k_{r_3} = k_r$,

$$k_r([\mathrm{H}] + [\mathrm{D}])^2 = k_{d_1}[\mathrm{H}_2] + k_{d_2}[\mathrm{D}_2], \tag{32}$$

since [HD] = 0 at the beginning of the reaction. If we introduce

$$[\mathrm{H}]_0 = \sqrt{\mathrm{K}_{\mathrm{H}_2}[\mathrm{H}_2]} = \sqrt{\frac{k_{d_1}}{k_r}[\mathrm{H}_2]} \text{ and } [\mathrm{D}]_0 = \sqrt{\mathrm{K}_{\mathrm{D}_2}[\mathrm{D}_2]} = \sqrt{\frac{k_{d_2}}{k_r}[\mathrm{D}_2]},$$

i.e., the thermodynamical [H] and [D] concentrations in pure H_2 and D_2 respectively, we find

$$[\mathrm{H}] + [\mathrm{D}] = \sqrt{[\mathrm{H}]_0^2 + [\mathrm{D}]_0^2}. \tag{32A}$$

The ratio [H]/[D] can be obtained from equation (29) by assuming for every phase of the reaction an establishment of an [H] and [D] concentration which remains stationary until the concentrations of H_2, HD and D_2 are unchanged.* Neglecting the terms accounting for dissociation and recombination we obtain

$$\frac{[\mathrm{H}]}{[\mathrm{D}]} = \frac{k_3[\mathrm{H}_2] + k_6[\mathrm{HD}]}{k_4[\mathrm{D}_2] + k_5[\mathrm{HD}]}, \tag{33}$$

or since $[\mathrm{H}_2] = [\mathrm{D}_2] = [\mathrm{H}_2]_0 - [\mathrm{HD}]/2$ and $k_3/k_5 = 2 \cdot 6$, $k_4/k_6 = 1 \cdot 5$ (the latter two ratios are over the temperature range investigated practically independent of the temperature)

$$\frac{[\mathrm{H}]}{[\mathrm{D}]} = \frac{k_3[\mathrm{H}_2]_0 + (0 \cdot 67\, k_4 - 0 \cdot 5\, k_3)[\mathrm{HD}]}{k_4[\mathrm{H}_2]_0 + (0 \cdot 38\, k_3 - 0 \cdot 5\, k_4)[\mathrm{HD}]}. \tag{33A}$$

It can be anticipated that k_3 will be of the same magnitude as k_4. For the special case that

$$k_4 < k_3 < 3k_4$$

the ratio [H]/[D] does not alter very much in the first half of the reaction ($[\mathrm{HD}] < [\mathrm{H}_2]_0/2$) and we can put approximately

$$\frac{[\mathrm{H}]}{[\mathrm{D}]} = \frac{k_3}{k_4}, \tag{33B}$$

---

* *Cf.* Bodenstein, 'Ann. Physik,' vol. 82, p. 836 (1927)

i.e., we can assume that [H] and [D] remain stationary even if $[H_2]$, $[D_2]$ and [HD] change. Introducing the atom concentrations

$$[\mathrm{H}] = \frac{k_3}{k_3 + k_4}\sqrt{[\mathrm{H}]_0^2 + [\mathrm{D}]_0^2} \quad \text{and} \quad [\mathrm{D}] = \frac{k_4}{k_3 + k_4}\sqrt{[\mathrm{H}]_0^2 + [\mathrm{D}]_0^2}$$

into equation (29) we obtain (again neglecting the terms accounting for recombination and dissociation)

$$\frac{d\,[\mathrm{HD}]}{dt} = \left\{\frac{\sqrt{[\mathrm{H}]_0^2 + [\mathrm{D}]_0^2}}{k_3 + k_4}\right\}\left\{2k_3 k_4\left([\mathrm{H}_2]_0 - \frac{[\mathrm{HD}]}{2}\right) - (k_3k_5 + k_4k_6)\,[\mathrm{HD}]\right\}. \quad (34)$$

Since in the mixtures used $[H_2] = [D_2]$ the equilibrium $H_2 + D_2 = 2HD$ is established when $[HD]^2/[H_2]\,[D_2] \simeq 4$ or when $[HD] = [H_2]_0$. Therefore we have

$$\frac{d\,[\mathrm{HD}]}{dt} = \left\{\frac{\sqrt{[\mathrm{H}]_0^2 + [\mathrm{D}]_0^2}}{k_3 + k_4}\right\} 2k_3k_4\,\{[\mathrm{H}_2]_0 - [\mathrm{HD}]\}, \quad (35)$$

or integrated

$$[\mathrm{H}_2]_0 - [\mathrm{HD}] = [\mathrm{H}_2]_0\, e^{-\{\sqrt{[\mathrm{H}]_0^2 + [\mathrm{D}]_0^2}\}\frac{2k_3k_4}{k_3+k_4}t}. \quad (36)$$

That the condition of the stationary H and D concentration is fulfilled is shown by the fact that the formula for the progress of the reaction derived from the kinetic measurements has a similar form. Experimentally it was found that the interaction $H_2 + D_2 = 2DH$ proceeds at 950° K 1·5 times more slowly than the reaction $H + H_2 \rightleftharpoons H_2 + H$.

This means that

$$\frac{k_1\,[\mathrm{H}]}{1{\cdot}5} = 2\,\frac{k_3k_4}{k_3 + k_4}\,\{\sqrt{[\mathrm{H}]_0^2 + [\mathrm{D}]_0^2}\}, \quad (37)$$

where k_1 denotes the velocity constant and [H] the atom concentration in the reaction $H + H_2 \rightleftharpoons H_2 + H$.

The ratio $[\mathrm{H}]/\sqrt{[\mathrm{H}]_0^2 + [\mathrm{D}]_0^2}$ is given by

$$\frac{[\mathrm{H}]}{\sqrt{[\mathrm{H}]_0^2 + [\mathrm{D}]_0^2}} = \frac{\sqrt{2K_{\mathrm{H}_2}}}{\sqrt{K_{\mathrm{H}_2} + K_{\mathrm{D}_2}}}, \quad (38)$$

where K_{H_2} and K_{D_2} denote the dissociation constants of H_2 and D_2 respectively and the factor $\sqrt{2}$ arises, since in a 50% H_2 + 50% D_2 mixture the concentration of H_2 and D_2 is half of that of H_2 in pure hydrogen at the same pressure.

At 950° K $[\mathrm{H}]/\sqrt{[\mathrm{H}]_0^2 + [\mathrm{D}]_0^2} = 1{\cdot}14$ and thus we obtain

$$\frac{k_3k_4}{k_3 + k_4} = 0{\cdot}38\,k_1. \quad (39)$$

Although it is not possible to deduce from this single equation both velocity constants k_3 and k_4 and the activation energies involved, one can estimate these values approximately in the following way.

The ratio of the velocity constants k_3 and the velocity constant k_1 for the reaction $\mathrm{H} + \mathrm{H}_2 = \mathrm{H}_2 + \mathrm{H}$ is given by

$$\frac{k_3}{k_1} = \sqrt{\frac{(\mathrm{M}_2 + \mathrm{M}_{11})\,(\mathrm{M}_1 \,.\, \mathrm{M}_{11})}{(\mathrm{M}_2 \,.\, \mathrm{M}_{11})\,(\mathrm{M}_1 + \mathrm{M}_{11})}}\; e^{-(\mathrm{W}_3 - \mathrm{W}_1)/\mathrm{RT}} = 0{\cdot}82\, e^{-(\mathrm{W}_3 - \mathrm{W}_1)/\mathrm{RT}}, \quad (40)$$

where the expressions under the square root account for the different collision frequencies in the reactions (3) and (1) and W_3 and W_1 denote the energies of activation of the reaction (3) and (1) respectively. Since it is to be expected that W_3 will be lowered by a few hundred calories compared with W_1 (on account of the smaller zero-point energy of the activated stated HHD compared with HHH, *see below*), we may take $k_3 = k_1$ for 950° K. For this value we find from equation (39) $k_4 = 0{\cdot}61 k_3$ or $k_3/k_4 = 1{\cdot}63$ (according to Table VII $k_3 = 1{\cdot}14 k_1$ and therefore $k_3/k_4 = 2{\cdot}0$).

We can obtain independent rough estimate of the different velocity constants and energies of activation by considering the zero-point energies of the activated complexes, and it will be seen that they are in satisfactory agreement with those deduced from experiment.

In general the vibrational frequencies of a triatomic molecule are given by[*][†] (*cf.* formulae (21) and (22))

$$p_1^2 + p_2^2 = f\,(1/\mathrm{M}_1 + 2\mathrm{M}/_2 + 1/\mathrm{M}_3),$$
$$p_1^2 \,.\, p_2^2 = f^2\,(1/\mathrm{M}_1\mathrm{M}_2 + 1/\mathrm{M}_2\mathrm{M}_3 + 1/\mathrm{M}_3\mathrm{M}_1),$$

M_2 denoting the atomic weight of the central atom and M_1 and M_3 the atomic weights of the two end atoms in the molecule (p_1 and p_2, *cf.* formulae (21) and (22)). Using these relations we are able to deduce the frequencies for the complexes

HHH	......	$\mathrm{M}_1 : \mathrm{M}_2 : \mathrm{M}_3 = 1:1:1$
HHD	......	$\mathrm{M}_1 : \mathrm{M}_2 : \mathrm{M}_3 = 1:1:2$
HDH	......	$\mathrm{M}_1 : \mathrm{M}_2 : \mathrm{M}_3 = 1:2:1$
HDD	......	$\mathrm{M}_1 : \mathrm{M}_2 : \mathrm{M}_3 = 1:2:2$
DHD	......	$\mathrm{M}_1 : \mathrm{M}_2 : \mathrm{M}_3 = 2:1:2$
DDD	......	$\mathrm{M}_1 : \mathrm{M}_2 : \mathrm{M}_3 = 2:2:2$

---

[*] Bailey and Cassie, 'Proc. Roy. Soc.,' A, vol. 135, p. 375 (1932).

[†] The authors have to thank Dr. G. B. B. M. Sutherland for this reference and discussions.

The zero-point energy will be then proportional to p_2 if we consider again only the symmetrical vibration (*cf.* p. 136) and neglect the deformation vibration (Table VI).

TABLE VI

	$\frac{p_1^2 + p_2^2}{f}$	$\frac{p_1^2 \times p_2^2}{f^2}$	$p_1/\sqrt{f}$	$p_2/\sqrt{f}$	E° in k cals
HHH	4·0	3·0	1·73	1·00	4·37
HHD	3·5	2·0	1·67	0·85	3·72
HDH	3·0	2·0	1·41	1·00	4·37
DHD	3·0	1·25	1·58	0·707	3·09
HDD	2·5	1·25	1·34	0·83	3·63
DDD	2·0	0·75	1·22	0·707	3·09

According to these calculations the energy of activation of the reaction $H + D_2 \rightleftharpoons HD + D$ is larger by 1·7 kg cals than that of the reaction $D + H_2 \rightleftharpoons HD + H$, which is nearly the same energy difference as that corresponding to the experimental value of k_3/k_4.

It must be emphasized once more that the zero-point energies of the activated state calculated in this manner are not necessarily correct, since the potential surface is different for these complexes when compared with that for ordinary molecules for which the relations are valid. It is to be expected, however, that the differences in the vibration frequencies, and consequently the differences in the zero-point energies, for the different activated complexes HHH, HHD, etc., are of the right order of magnitude. The complete system of energy levels occurring in the interaction of hydrogen and deuterium is shown in fig. 7. Table VII summarizes the results of the calculation and shows the zero-point energies in the initial and activated state, the energies of activation, and the relative collision numbers and relative velocity constants for the individual reactions involving one atom and one molecule.

The velocity constants for the reactions 3–8 were worked out according to formulae similar to (40),* assuming for the zero-point energies of the activated complexes the values calculated and listed in Table VI. It should be pointed out that the velocity constants k_3 and k_4 appearing in the interaction of H_2 and D_2 calculated in this manner are in exact agreement with the experiments giving for

$$\frac{k_3 k_4}{k_3 + k_4} = \frac{1 \cdot 14 \times 0 \cdot 54 k_1}{1 \cdot 14 + 0 \cdot 54} = 0 \cdot 376 k_1$$

compared with the experimental figure of $0 \cdot 38 k_1$ (*cf.* formula (39), p. 143).

---

* [*Note added in proof, July* 10, 1935—The data of Table VII, calculated according to classical theory, have to be considered preliminary. An exact re-calculation of these data is to be published in the paper of Farkas and Wigner.]

The reactions (5) and (7) are the inverse reactions to (3) and (4) respectively. The reactions (6) and (8) are entirely symmetrical and do not involve a change in the structure of the molecule. The velocity constants for the reactions (5), (6), (7), and (8) are only half of what one would have expected on the basis of a formula similar to formula (40),

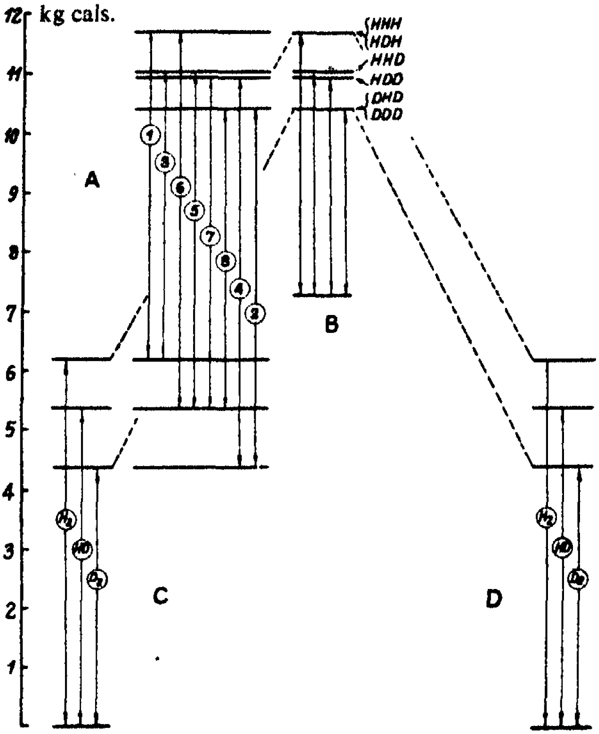


FIG. 7—The numbers for the energies of activation refer to the numbers of the reactions given in Table VII. A, energies of activation; B, zero-point energies in the activated state; C, zero-point energies in the initial state; D, zero-point energies in the final state.

since the atom attacking the HD molecule is able to undergo the alternative reactions:

$$H + \begin{cases} DH = HD + H \\ HD = H_2 + D \end{cases} \qquad (42)$$

$$D + \begin{cases} HD = DH + D \\ DH = D_2 + H \end{cases}, \qquad (43)$$

depending on which side of the molecule it strikes. It is worth while noting that by replacing a hydrogen atom by a deuterium atom the energy

TABLE VII

Reaction	Initial state	Initial zero-point energy (in k cals)	Activated state	Zero-point energy (in k cals) in the activated state	Relative collision number	Energy of activation in k cals	Relative velocity constant at 1000°K
1. $H + H_2 \rightleftharpoons H_2 + H$	$H + H_2$	6·18	HHH	4·37	1·00	5·5	k_1 1·00
1A. $H + o\text{-}H_2 = p\text{-}H_2 + H$	$H + H_2$	6·18	HHH	4·37	1·00	5·5	k'_1 0·25
1B. $H + p\text{-}H_2 = o\text{-}H_2 + H$	$H + H_2$	6·18	HHH	4·37	1·00	5·5	k''_1 0·75
2. $D + D_2 \rightleftharpoons D_2 + D$	$D + D_2$	4·39	DDD	3·09	0·71	5·5 + 0·51	k_2 0·55
2A. $D + o\text{-}D_2 = p\text{-}D_2 + D$	$D + D_2$	4·39	DDD	3·09	0·71	5·5 + 0·51	k'_2 0·18
2B. $D + p\text{-}D_2 = o\text{-}D_2 + D$	$D + D_2$	4·39	DDD	3·09	0·71	5·5 + 0·51	k''_2 0·37
3. $D + H_2 = HD + H$	$D + H_2$	6·18	HHD	3·72	0·82	5·5 − 0·65	k_3 1·14
4. $H + D_2 = HD + D$	$H + D_2$	4·39	HDD	3·63	0·92	5·5 + 1·05	k_4 0·54
5. $H + HD = H_2 + D$	$H + HD$	5·36	HHD	3·72	0·94	5·5 + 0·17	k_5 0·43
6. $D + DH = D_2 + H$	$D + DH$	5·36	HDD	3·63	0·75	5·5 + 0·08	k_6 0·36
7. $H + DH = HD + H$	$H + HD$	5·36	HDH	4·37	0·94	5·5 + 0·82	k_7 0·31
8. $D + HD = HD + D$	$D + HD$	5·36	DHD	3·09	0·75	5·5 − 0·46	k_8 0·47

of activation appears to be decreased, whereas the replacement of a hydrogen molecule by a deuterium molecule causes the energy of activation to be increased, as was predicted by Polanyi.*

Finally it is interesting to consider the effect of the zero-point energy in the activated state in reactions of the type

$$X + H_2 = XH + H \tag{44}$$

$$X + D_2 = XD + D \tag{44A}$$

and

$$H + X_2 = HX + X \tag{45}$$

$$D + X_2 = DX + X, \tag{45A}$$

where X designates a relatively heavy element.

Reactions of this type occur, for example, in the union of hydrogen and chlorine, according to the well-known chain mechanism

$$Cl + H_2 = HCl + H \tag{46}$$

$$H + Cl_2 = HCl + Cl, \tag{47}$$

and in the union of deuterium and chlorine

$$Cl + D_2 = DCl + D \tag{46A}$$

$$D + Cl_2 = DCl + Cl. \tag{47A}$$

We shall estimate the zero-point energies of the activated complexes involved in these particular reactions according to the formula (48), assuming that the force-constant for the hydrogen-chlorine bond is the same as that for the hydrogen-hydrogen bond and the force-constant for the chlorine-chlorine bond is half of this value. Table VIII shows the zero-point energies obtained for the different complexes by considering only the symmetrical vibrations. It will be recognized that as the total mass of the activated complex increases, its zero-point energy decreases

$$\left.\begin{aligned} p_1^2 + p_2^2 &= f(1/M_1 + 1/M_2) + f'(1/M_2 + 1/M_3) \\ p_1^2 \cdot p_2^2 &= ff'(1/M_1M_2 + 1/M_2M_3 + 1/M_3M_1) \end{aligned}\right\} \tag{48}$$

(f and f' being the force constants corresponding to the bond between the atoms M_1-M_2 and M_2-M_3 respectively),

and, consequently, so does the separation between the energy levels of the activated complexes containing different isotopes.

---

* 'Nature,' vol. 133, p. 26 (1934).

Thus the separation between the energy levels of HHH and DDD is 1·3 kg cals, between ClHH and ClDD 0·8 k cals, and between HClCl and DClCl practically nothing. Consequently the influence of the zero-point energy in the activated state on the energy of activation is the smaller the larger the mass of the activated complex. We can claim no high order of accuracy in view of the rather rough approximation made in this calculation, but a comparison with experiments shows that they are qualitatively correct.

TABLE VIII

Initial state	Initial zero-point energy in k cals	Activated state	Zero-point energy in k cals in the activated state
$Cl + H_2$	6·2	ClHH	2·8
Cl + DH	5·4	ClDH	2·5
$Cl + D_2$	4·4	ClDD	2·0
$H + Cl_2$	0·8	HClCl	0·7
$D + Cl_2$	0·8	DClCl	0·7

According to measurements of Rollefson* and L. and A. Farkas,† the ratio of the velocities for the reactions

$$Cl + H_2,\ Cl + HD,\ \text{and}\ Cl + D_2\ \text{is at 30–32° C,}$$

1 : 0·3 : 0·1. If we allow for the different collision numbers, we can conclude that the energies of activation of the reactions

$$Cl + HD = DCl + H \qquad \text{and} \qquad Cl + D_2 = DCl + D$$

are 0·6 and 1·2 k cals higher respectively than that of the reaction

$$Cl + H_2 = HCl + H.$$

Since the difference in the zero-point energies of the molecules H_2, HD, and D_2 is 0·8 and 1·8 k cals respectively, we have to assume that the level ClHH is separated from the levels ClDH and ClDD by 0·2 and 0·6 kg cal respectively in satisfactory agreement with the values given in Table VIII.

According to Bach, Bonhoeffer, and Moelwyn-Hughes, the difference in the activation energies of the reactions

$$H_2 + Br = HBr + H$$

---

* 'J. Chem. Phys.,' vol. 2, p. 144 (1934).

† 'Naturwiss.,' vol. 22, p. 218 (1934).

and

$$D_2 + Br = DBr + D$$

is about 1·5 k cals, *i.e.*, only 0·3 k cal smaller than the difference in the zero-point energies of the initial state, and again this discrepancy (if at all real) might be explained by a separation in the levels of the activated complexes involved.

A conclusive evidence for the equal reactivity of H and D atoms in a reaction of the type

$$H + X_2 = HX + X$$

$$D + X_2 = DX + X$$

has not yet been found; there are, however, some experiments which seem to prove this indirectly.

According to Melville,* the rate-determining step in the mercury-sensitized photochemical interaction of hydrogen and nitrous oxide is at 300° C and upwards,

$$H + N_2O = HO + N_2.$$

The substitution of deuterium for hydrogen does not alter the rate of this interaction, showing that the reaction involving D-atoms has the same energy of activation as that involving H-atoms.

The authors are greatly indebted to Professor E. K. Rideal, F.R.S., for his constant interest in this investigation and for providing all facilities necessary for the present experiments. They have also to thank Dr. O. H. Wansbrough-Jones for reading the manuscript, and the Central British Fund for German Jewry for financial assistance.

SUMMARY

It is shown that in the temperature region 850°–1000° K and at pressures 3 to 74 mm Hg both the reconversion of ortho-deuterium and the interaction of hydrogen and deuterium proceed in the homogeneous gas phase according to the atomic mechanisms

$$D + o\text{-}D_2 \rightleftharpoons p\text{-}D_2 + D \quad \text{and} \quad \begin{array}{l} D + H_2 \rightleftharpoons HD + H \\ H + D_2 \rightleftharpoons HD + D, \end{array}$$

respectively.

---

* Melville, 'J. Chem. Soc.,' p. 1243 (1934).

The rates of these reactions have been compared with that of the thermal para-hydrogen conversion, and it was found that the reconversion of ortho-deuterium proceeds 2·4 times and the interaction of hydrogen and deuterium 1·5 times more slowly than the reconversion of para-hydrogen at the same temperature and pressure.

In the ortho-deuterium conversion the different rate is mainly due to the different collision frequency and concentration of deuterium atoms, and the energy of the activation for the reaction $D + D_2 \rightleftharpoons D_2 + D$ is only 0·5 k cal higher than that of the reaction $H + H_2 \rightleftharpoons H_2 + H$. On the other hand, the different reaction rate in the interaction of hydrogen and deuterium is caused by the fact that the energy of activation for the partial reaction $H + D_2 \rightleftharpoons HD + D$ is higher by about 1·7 k cals than that of the reaction $D + H_2 \rightleftharpoons HD + H$.

This behaviour is explained by assuming that the activated complexes HHH, DHH, HDD, and DDD possess zero-point energies similar to the molecules H_2, HD, and D_2, and it is shown that by considering the possible vibrations in the activated complexes, a scheme for the energy levels involved in these interchanges can be worked out which is consistent with all experimental results.

Experiments on Heavy Hydrogen*

VI—The Ratio of the Magnetic Moment of the Proton to the Magnetic Moment of the Deuteron

By ADALBERT FARKAS and LADISLAS FARKAS, Department of Colloid Science, The University, Cambridge

(*Communicated by E. K. Rideal, F.R.S.—Received April* 5, 1935)

It has been pointed out† that it is possible to determine the ratio of the magnetic moment of the proton to that of the deuteron by comparing the rates of the ortho-para interconversion for hydrogen and deuterium which proceed in the presence of oxygen.

The catalytic ortho-para-hydrogen-conversion induced by oxygen was discovered by L. Farkas and Sachsse,‡ and is caused by the inhomogeneous magnetic field of the para-magnetic oxygen molecule interacting during the collision with the nuclear magnets of the hydrogen molecule.

The reaction proceeds according to the equation

$$(p_t - p_\infty) = (p_0 - p_\infty)\, e^{-k[O_2]}\, t \qquad (1)$$

where p_t p_0 and p_∞ denote the percentage of para-hydrogen at the time t, $t = 0$ and $t = \infty$, k the velocity constant, and $[O_2]$ the concentration of oxygen. k was found to increase slightly with increasing temperature.

This magnetic catalysis of the ortho-para-conversion was first examined theoretically by Wigner.§ He showed that the probability of an ortho-para transition in a collision with a para-magnetic molecule like oxygen is *ceteris paribus* proportional to the square of the magnetic moment of the proton and to the square of the magnetic moment of the collision partner, and that the dependence of the rate of the conversion on temperature could be theoretically accounted for in a satisfactory way.

A more refined theory of this conversion was developed by Kalckar and Teller,|| especially in respect to the determination of the ratio of the

---

* *Cf.* L. and A. Farkas, 'Nature,' vol. 135, p. 372 (1935).

† A. and L. Farkas and P. Harteck, 'Proc. Roy. Soc.,' A, vol. 144, p. 481 (1934).

‡ 'S. B. preuss. Akad. Wiss.,' p. 268 (1933); 'Z. phys. Chem.,' B, vol. 23, pp. 1, 19 (1933).

§ 'Z. phys. Chem.,' B, vol. 23, p. 28 (1933).

|| Kalckar and Teller, 'Nature,' 134, 180 (1934); and 'Proc. Roy Soc.,' A., vol. 150, p. 520 (1935).

magnetic moments of the proton and deuteron (μ_H/μ_D). They showed that the different mechanical moments of the proton and deuteron have an influence on the transition probabilities and pointed out that the ratio of the magnetic moments of the proton and deuteron could be obtained with greater reliability if the velocities of the corresponding conversions were compared at the absolute temperature 2T for H_2 and T for D_2. In this case the following formula is obtained connecting the velocity constants (k), mechanical moments (S), and magnetic moments (μ):

$$\frac{k_{H_2}^{(2T)}}{k_{D_2}^{(T)}} = \frac{\mu_H^2\, S_D^2\, (2\, S_H + 1)\, [\zeta\, (S_H + 1) + S_H]}{\mu_D^2\, S_H^2\, (2\, S_D + 1)\, [S_D + 1 + \zeta\, S_D]}, \tag{2}$$

where ζ is defined by

$$\zeta = \frac{\sum\limits_{J=\text{odd}} (2\,J + 1)\, e^{-J(J+1)h^2/8\pi^2IkT}}{\sum\limits_{J=\text{even}} (2\,J + 1)\, e^{-J(J+1)h^2/8\pi^2IkT}}, \tag{3}$$

where J = rotational quantum number and I = moment of inertia.

3ζ is the ratio of the equilibrium concentration of ortho-hydrogen and of para-hydrogen at the temperature 2T or $\zeta/2$ is the ratio of the equilibrium concentration of para-deuterium and ortho-deuterium at the temperature T.

With $S_D = 1$, $S_H = \frac{1}{2}$, the formula (2) assumes at higher temperatures ($T > 120°K$), ζ being unity, the following very simple form

$$\frac{k_{H_2}^{(2T)}}{k_{D_2}^{(T)}} = \frac{8}{9}\left(\frac{\mu_H}{\mu_D}\right)^2. \tag{4}$$

Apart from its formal simplicity this formula has the advantage that it can be derived without making any special assumption about the mechanism of collision and by neglecting only the motion of the oxygen molecules, the effect of the slight van der Waals forces and magnetic moments other than dipole moments.

The subject of the present paper is the determination of the ratio of the magnetic moments of the proton and of the deuteron by comparing the rates of the ortho-para-conversion for hydrogen and deuterium at corresponding temperatures. It may be stated at once that over a fairly large range of temperature the ratio of the magnetic moments of the proton and deuteron was found to vary only slightly, and thus the present experiments give an accurate value for this ratio and confirm at the same time the theory proposed by Kalckar and Teller.

The ortho-deuterium was prepared by adsorption of pure deuterium on charcoal at the temperature of liquid hydrogen and by subsequent desorption.* At 20° K the equilibrium mixture contains 98% o-D_2 + 2% p-D_2. The ortho-deuterium was then mixed with a certain amount of oxygen and the gradual reconversion investigated by extracting small samples of gas of 0·003 cc and analysing its o-D_2 content by means of the micro-conductivity method.* The oxygen present in the hydrogen was frozen out in a trap cooled by liquid hydrogen before the sample was introduced into the conductivity cell.

In previous experiments* it has been found that traces of solid oxygen are able to catalyse the formulation of new o-D_2 in the partly converted gas when passing the trap at 20° K.† For this reason particular care was taken to use a trap of suitable dimensions which would permit the passage of the gas with a sufficient speed when the catalytic conversion of deuterium is negligible and does not affect the accuracy of the measurements. This was tested in special experiments by passing normal D_2 through the trap and ascertaining by analysis whether the o-D_2 content had changed.

Three series of experiments were carried out with ortho-deuterium at 293, 193, and 83° K, and also one run with para-hydrogen at 293° K in order to check the experimental arrangement.

The progress of the conversion was found to proceed according to the formula (1) and the velocity constant k was calculated from the relation

$$k = \frac{1}{t\,[O_2]}\, ln\, \frac{u_0}{u_t}, \tag{5}$$

where $u_0 = p_0 - p_\infty$ and $u_t = p_t - p_\infty$ denote the excess concentration of parahydrogen relative to equilibrium, t is given in minutes, and the concentration of O_2 is mole/litre. Table I shows the results with hydrogen at 20° C. The velocity constant of 8·34 (mole/litre min)$^{-1}$ is in satisfactory agreement with the original figure of Farkas and Sachsse 8·8 (mole/litre min)$^{-1}$‡

The results of an experiment at 20° C with ortho-deuterium are given in Table II, the constants being calculated according to a similar formula as (5), u denoting in this case the excess concentration of ortho-deuterium. The velocity constant obtained in four other runs at 293, 193, and 82° K are summarized in Table III, each velocity constant being the mean value of 6 or 7 constants calculated as shown in Table II.

* A. and L. Farkas and Harteck, 'Proc. Roy. Soc.,' A, vol. 144, p. 481 (1933), and part V.

† *Cf.* also Cremer, 'Z. phys. Chem.,' B, vol. 28, p. 199 (1935).

‡ Owing to an error in the numerical calculation in the original paper, this constant is given as 9·2 (mole/litre min)$^{-1}$.

TABLE I

p–H_2 = 29·5 mm Hg. O_2 = 43·5 mm Hg. T = 293° K. $[O_2]$ = 2·38 × 10^{-3} mole/litre

Time in minutes	litre u_t/u_0	k_{H_2}
0	1·000	—
2·5	0·950	8·54
8	0·855	8·22
15	0·750	8·09
29	0·579	7·89
37	0·478	8·44
47	0·391	8·43
57	0·312	8·61

mean 8·34 (mole/litre min)$^{-1}$

TABLE II

o-D_2 = 40 mm Hg. O_2 = 272 mm Hg T = 293° K. $[O_2]$ = 14·9 × 10^{-3} mole/litre.

Time in minutes	u_t/u_0	k_{D_2}
0	1·000	—
10	0·878	0·880
27	0·721	0·814
39	0·664	0·711
49	0·591	0·722
63	0·426	0·913
81	0·336	0·904
106	0·262	0·849

mean 0·828 (mole/litre min)$^{-1}$

TABLE III

Temperature °K	Concentration of oxygen in mole/litre	Velocity constant in (mole/litre min)$^{-1}$
293	8·61 × 10^{-3}	0·844
293	10·88 × 10^{-3}	0·819
293	14·9 × 10^{-3}	0·828
193	23·2 × 10^{-3}	0·696
83	38·8 × 10^{-3}	0·504

Table IV shows the comparison of these velocity constants with the corresponding velocity constants for the para-hydrogen conversion. The latter values are taken from the paper by L. Farkas and Sachsse.* The ratio of the magnetic moments μ_H/μ_D was calculated according to the formula (4) for the temperatures 293 and 193° and according to the formula

$$\frac{k_{H_2}^{(166)}}{k_D^{(83)}} = 1 \cdot 18 \left(\frac{\mu_H}{\mu_D}\right)^2,$$

which follows directly from formula (2), considering that at 83° K $\zeta = 0 \cdot 88$.

TABLE IV

Temperature (T) °K	$k_{D_2}^{(T)}$	$k_{H_2}^{(2T)}$	$\frac{k_{H_2}^{(2T)}}{k_{D_2}^{(T)}}$	$\frac{\mu_H}{\mu_D}$
83	0·504	6·3	12·5	3·85
193	0·696	10·1	14·5	4·03
293	0·830	12·3	14·8	4·07
				mean 3·96 ± 0·11

The slight variation of the ratio μ_H/μ_D with temperature is within the limits of possible experimental error, the limits being about 10% in the determination of the velocity constants and therefore about 5% in the determination of the ratio.

It is perhaps worth while to point out that qualitatively we might, on the basis of the theory developed by Kalckar and Teller, anticipate an apparent variation of the ratio μ_H/μ_D with temperature corresponding to the consequence of neglecting the van der Waals forces. Apart from this slight variation of the ratio μ_H/μ_D in the wide temperature range of 83° to 293° K, the results of this investigation may be regarded as a direct experimental proof for the theory of Kalckar and Teller, and consequently the ratio of the magnetic moments of the proton and of the deuteron obtained in the present experiments of 3·96 ± 0·11 is supported theoretically.

The ratio μ_H/μ_D obtained in the present paper can be compared with that deduced from recent measurements of the magnetic moments of the proton and of the deuteron by two different magnetic deflection methods, Table V.

On the basis of the present investigation it is not possible to decide which value is the better for μ_H and μ_D respectively, since the ratio μ_H/μ_D

---

* 'S. B. preuss. Akad. Wiss.,' p. 268 (1933); 'Z. phys. Chem.,' B, vol. 23, pp. 1, 19 (1933).

TABLE V

	μ_H	μ_D	μ_H/μ_D
Frisch and Stern, Estermann and Stern	2·5 ± 0·25	0·7	3·6
Rabi, Kellog, and Zacharias	3·25 ± 0·32	0·75 ± 0·2	4·3
This paper	—	—	3·96 ± 0·11

obtained in this investigation lies well between the limits of error involved in the magnetic deflection measurements.

The authors are very much indebted to Professor E. K. Rideal, F.R.S., for his hospitality and the facilities provided in his laboratory. Their thanks are also due to the Central British Fund for German Jewry for financial assistance.

SUMMARY

The rate of reconversion of ortho-deuterium to normal deuterium in the presence of oxygen has been measured over the temperature range 83° – 293° K. The velocity constants obtained have been compared with the corresponding velocity constants of the para-hydrogen conversion, and from the ratio of the velocity constants, on the basis of the theory developed by Kalckar and Teller, the ratio of the magnetic moments of the proton and of that of the deuteron have been determined as 3·96 ± 0·11.

On Chemical Reaction in the Electric Discharge
I—The Chemical Effects of Impulse Discharges

By E. J. B. Willey, Davy-Faraday Laboratory of the Royal Institution and Imperial Chemical Industries, Ltd.

(Communicated by F. A. Freeth, F.R.S.—Received April 6, *Revised July* 15, 1935)

Although a very large number of studies has been made upon the chemical reactions which can be induced by electric discharges, the effects of controlled condenser sparks appear to have been neglected, with one exception.

Wrede* has shown that if a large condenser (8 mfd) is charged to a high potential, and then allowed to discharge through a gas at low pressure, a spark gap being inserted in series with the discharge tube proper to increase the breakdown voltage, then hydrogen, oxygen, and nitrogen may be dissociated to the extent of 80, 60, and 40% respectively, the atom concentration being measured upon a most ingenious "diffusion gauge." The Wrede discharge has been extensively employed by Steiner and his associates in a series of semi-spectroscopic investigations upon active nitrogen and atomic hydrogen,† but its use as a means of effecting chemical reaction has not so far received attention.

In the following communication it will be shown that Wrede's results for nitrogen may be confirmed by chemical means and by simpler apparatus, and that the controlled impulse discharge is very superior, both as regards absolute and power yields as compared with A.C. and D.C. discharges.

The following considerations must be borne in mind.

1. A condenser of capacity C farads charged by a mean current of i_c amperes for t seconds will attain a voltage V given by

$$V_p = i_c t/C. \qquad (1)$$

If this voltage be that for the breakdown of a discharge tube connected in parallel with the condenser, the potential will then fall abruptly to the breakoff value V_b, when the cycle of charging and discharging will recommence. A rhythmic discharge will thus take place, whose duration (of the order of microseconds) will be very small compared with the interval between pulses, provided that the charging current is not too high and that the inductance in circuit is small. Its frequency f will equal the reciprocal of t in the equation—

---

* 'Z. Physik,' vol. 53, p. 54 (1929).

† *E.g.*, Bay and Steiner, 'Z. phys. Chem.,' B, vol. 9, p. 93 (1930); Steiner, 'Z. Elecktrochem.,' vol. 36, p. 807 (1930).

$$(V_p - V_b) = i_c t/C, \text{ i.e., } \quad f = i_c/C\,(V_p - V_b). \tag{2}$$

2. When the condenser discharges, the energy expended in the circuit J is given by—

$$J = \tfrac{1}{2}C\,(V_p^2 - V_b^2) \text{ joules } (= \text{watt-seconds}) \text{ pe1 flash,} \tag{3}$$

and the power W equals J . f watts, *independently of both duration and wave form of the current and valtage.* In gases at low pressures, where V_b is of the order of 0·1 V_p, we may write—

$$W = \tfrac{1}{2}fCV_p^2 \tag{4}$$

without introducing a large error. At higher pressures, the approximation is no longer permissible, since V_b approaches V_p in magnitude.

It is clear that these simple considerations are valid only when the rate of charging of the condenser is such that no appreciable restoration of the initial voltage can occur during the period of the discharge; if i_c is so great that the condenser voltage never falls to V_b the discharge will no longer be intermittent but continuous, and its characteristics will entirely alter. Moreover, when, as in the present experiments, half-wave rectified A.C. is employed to charge the condenser, the maximum frequency of a true rhythmic impulse discharge cannot exceed that of the A.C. supply (50 cycles). Under these conditions, the condenser charges to the breakdown voltage and discharges once in every positive half-cycle, and other conditions have therefore not been admitted.

Consider now the case of a gas streaming through a narrow tube in which a periodic flashing is maintained by a properly adjusted Wrede circuit. If the flow speed be correctly regulated with respect to the discharge frequency, the gas can be acted on by a single pulse in the time it spends in the tube, the energy applied being that stored up by the condenser in the interval between the flashes, which may be regarded as the time taken by a molecule to travel, under the influence of the gas flow, from one end of the tube to the other. Similarly, two or more pulses may occur in this time, but there will also be a minimum value of f below which a molecule may pass undisturbed through the tube; this critical frequency (f(min)), which is a constant for any tube under given conditions of gas flow and pressure, will be given by

$$f(\text{min}) = 760\,F/3600\,pv \tag{9}$$

where F = flow speed of the gas in cc/hr at S.T.P., p = gas pressure in mm mercury, v = volume of discharge tube in cc. If ever a tube is operated with $f < f$(min), the flow speed of the gas for purposes of calculation must be taken as F . f/f(min).

We may now consider some probable consequences of the action of an impulse discharge upon a gas.

(*a*) Since, as will be shown later, the instantaneous current in a condenser discharge through a gas rises to very high values (200 A has sometimes been observed by the author), it follows that the instantaneous concentration of positive ions must be increased from that prevailing in the A.C. discharge if we consider these quantities to be determined at a given time interval shortly after discharge has commenced; as the voltage across the condenser approaches the extinction potential of the discharge, so does the current fall and there will be little, if any, distinction between an impulse discharge at this point of its career and an ordinary A.C. discharge at the same stage, always provided, of course, that the former is aperiodic. Now investigations upon reactions taking place under the influence of controlled electrons and α-particles, etc., as well as those in D.C. glow discharges, have made it quite clear that under those conditions chemical change is in the greater majority of reactions preceded by the formation of positive ions, although what happens later depends upon the particular reaction and the experimental conditions ; one can say, however, that it is so complicated that only in very simple cases can any reasonably certain answer be given, and it is best to regard the positive ions as the parents of a large variety of other atomic and molecular species, many of which may possess abnormal chemical reactivity.* We may thus expect a fairly considerable enhancement of chemical reaction in unit time in passing from a D.C. or A.C. discharge to one of the impulse variety, although if we still have to deal with the simple law of electrochemical equivalence found by Keith Brewer and Finch and their associates to hold for a large number of reactions, viz., the amount of chemical change in unit time varies directly as the current flowing, one would expect the increased velocity due to the much larger currents to be offset by the much smaller duration of the discharge. This should not be so noticeable if a square law is operative. Whether or not this principle holds for impulse discharge reactions could only be established by determining the mean current through the gas and other discharge characteristics such as wave form. There does not, however, seem to be any reason to expect that the *products* of an impulse discharge reaction will be other than those where ordinary A.C. or D.C. is used, since the changes should be of degree rather than kind, except where certain controllable secondary effects become important (*see* later).

---

* *Cf.* Caress and Rideal, 'Proc. Roy. Soc.,' A, vol. 120, p. 370 (1928); vol. 115, p. 684 (1927). Wansbrough-Jones, *Ibid.*, vol. 127, p. 511 (1930). Numerous other references are given by Willey, 'Trans. Faraday Soc.,' vol. 30, p. 471 (1934).

(*b*) There is one important respect in which an impulse discharge reaction may differ from one in which A.C. or D.C. is concerned, *i.e.*, by reason of its very short duration it may give a larger proportion of the primary products if these be stable compounds which need further thermal or electrical treatment to be converted into the substances obtained with A.C. or D.C. An example of this will be quoted in a later communication, in which it is shown that the synthesis of hydrogen cyanide from nitrogen and methane in a discharge probably involves the intermediate formation of an entity (CH) which further reacts with nitrogen to give the product concerned.

The same considerations may hold if one or more of the primary products is a metastable body or a free radical whose life period exceeds the duration of the impulse which creates it, whether this genesis be a single- or multi-stage reaction; when A.C. or D.C. is in use, a large proportion of these species must undoubtedly be destroyed in secondary reactions involving excitation, etc. An example of this is the diminution of atom concentration in nitrogen which is subjected to the combined action of impulse and weak high-frequency discharges, as observed by Wrede (*loc. cit.*).

(*c*) It may be suggested that the ability of the impulse discharge to oscillate when $R^2/4L^2 < 1/LC$ (R being the resistance and L the inductance in circuit) might be turned to good account in some cases, through secondary excitation of the products of the initial pulse, especially since the oscillations tend to run at a high frequency. Evidence is not lacking, although very scanty, that the products and yields at 10^6–10^7 cycles differ from those at commercial frequencies; mention may be made of the interaction of hydrogen and carbon dioxide to give a water gas equilibrium* instead of the mixture of formaldehyde and formic acid obtained at 50 cycles.† The production of nitrogen afterglows in *air*, which is either impossible or very difficult with other discharges,‡ and the noticeable effect of frequency upon the ignition of carbon monoxide-oxygen mixtures§ may also to a certain extent be examples of the same phenomenon.

---

* Lunt, 'Proc. Roy. Soc.,' A, vol. 108, p. 173 (1925).

† Moser and Isgarischew, 'Z. Elektrochem.,' vol. 16, p. 613 (1910).

‡ Lewis, 'Phys. Rev.,' vol. 31, p. 314 (1928); 'J. Amer. Chem. Soc.,' vol. 51, pp. 654, 655 (1929); Herzberg, 'Z. Physik,' vol. 46, p. 878 (1928); Kaplan, 'Proc. Nat. Acad. Sci., Wash.,' vol. 14, p. 258 (1928); MacCullum and Perry, 'Nature,' vol. 121, p. 942 (1928).

§ Finch and Thompson, 'Proc. Roy. Soc.,' A, vol. 134, p. 343 (1931).

(*d*) It seems possible that the different conditions prevailing in the breakdown and main discharge periods might be used here. The time elapsing between commencement of the discharge and the establishment of steady conditions (*i.e.*, division of the discharge into well-marked zones) is only some 10^{-9} seconds, but it is clear *that in this time-interval circumstances later obtaining only in the cathode zone extend throughout the whole of the discharge.*

It follows, then, that if (1) the breakdown period is not too short compared with the total duration of the discharge, and (2) the development of the positive column does not lead to a destruction of products first formed in this induction period, the total effects may, in their chemical aspects, approximate to those found in the many cathodic reactions which have been studied. But as regards the *amounts* of reaction products attributable to the two periods concerned, we can only say that the ratio will depend upon the current passing in each case. This leads to the consequence that where a reaction proceeds through the formation of an excited state only, it would be best to employ a discharge of long breakdown period and with a total duration as short as possible; but where positive ions are concerned, the main discharge period would be of most importance, and this, as already mentioned, seems to hold in the majority of cases. These suggestions could easily be tested by working in the Townsend region of the discharge.

ELECTRICAL MEASUREMENTS

Measurement of the power applied to an impulse discharge involves the evaluation of f, V_p, and V_b, and they have severally been determined as follows:

Frequency—When the discharge is being operated at low pressures, f may be calculated from equation (2); but it is generally desirable to measure it directly. It is clear that provision of a frequency-measuring device—especially one of an inductive nature—in the actual discharge circuit might introduce difficulties, and, since the form of the current-voltage characteristic changes somewhat from case to case, a device accurate in one might read incorrectly in another. In these investigations, therefore, the periodicity has been determined by means of a photocell connected to an amplifier and recorder.

A simple two-valve circuit, with transformer coupling, works well, provided that it is designed to handle signals of large amplitude and short duration. The output is fed to a drum recorder (a modified " Fultograph " picture receiver), carrying a cylinder of moist paper impregnated with 5% potassium iodide. Since the diameter of the dots of iodine thus produced never exceeds that of the stylus (0·3 mm), it is to be concluded—the period of rotation of the drum being known—that the duration of each discharge is not more than 2×10^{-3} seconds.

Peak Voltages—These, together with the R.M.S. values, were measured upon the rectifier-condenser circuit* in conjunction with a non-inductive potential divider of 10 megohms shunted across the discharge. (Peak currents were similarly determined, when required, by measurement of the maximum voltage developed across a non-inductive resistance of 0·5 ohm connected in series with the discharge tube. In some cases the values obtained were as high as 200 amperes.) The mean voltage was determined by means of a moving-coil milliammeter connected in series with the potential divider. The circuits were checked for accuracy according to standard practice. By means of a reversing switch it was possible to interchange the voltmeter connexions to the tappings on the potential divider. If the discharge were truly aperiodic, one end of the resistance would always be positive with respect to the other, whereas this would not be so if the spark were oscillatory. No such reversal of polarity was ever found, in accordance with expectations from the short duration of the discharge, and it was therefore concluded that in these experiments a dead-beat current pulse was obtained. (Oscillograms taken in later experiments have confirmed this supposition.)

Break-off Voltages—When required, these were calculated from the difference of *f* as observed by means of the counter and as given by the reciprocal of *t* in equation (1). These results always agreed with those to be expected from the geometry of the discharge tube and the nature and pressure of the gas therein.

POWER MEASUREMENTS UPON OTHER DISCHARGES

Direct Current Discharges—In this case the power is accurately given by the volt-ampere product, provided that the discharge is free from superimposed oscillations which are partly plasmoidal in origin but also attributable, in some instances, to the liberation of gas from the electrodes.† It was found that when the tube was adequately water-cooled, no sound was audible in a loudspeaker connected to the photocell amplifier described above, and that a portable radio-receiver placed alongside the discharge gave no note, whereas, were oscillations present, a harmonic, if not a fundamental, should have been noticeable. With heavy currents, 200 mA, and a short spark, 7 cm, the discharges became unstable, a " flare " being produced which was accompanied by rapid disintegration of the electrodes, violent fluctuations of both current and voltages and very strong pick-up in both of the amplifiers; the same trouble was encountered with the A.C. discharge, and quantitative work was not attempted under such conditions.

---

* Davies, Standring, and Bowdler, 'J. Inst. Elect. Eng.,' vol. 68, p. 1228 (1930).

† Fox, 'Phys. Rev.,' vol. 35, p. 1066 (1930).

Alternating Current Discharges—The power may in this case be calculated with fair accuracy by a wattmeter connected in the primary side of the transformer, accompanied by a knowledge of the iron and copper losses and the secondary voltage as read upon an electrostatic instrument.* This involves the same proviso as to oscillations as is necessary with the D.C. discharge; but since it was found that the only note given by the amplifiers was that corresponding to the line frequency (50 cycles) with a fairly sharp cut-off on either side, no large inaccuracy in the measurements seems likely. The mean current through the discharge (read upon the usual arrangement of two diodes and two milliammeters) was directly proportional to the primary wattage from the lowest value which could be recorded, 5 mA, up to the point at which instability set in.

EXPERIMENTAL

Half-wave rectified A.C. was used in all the investigations with impulse discharges, an arrangement which enables the valve filament and one end of the discharge tube to be earthed, but limits f to not more than the line frequency (*cf.* pp. 159–160). The rate of charging of the condenser was controlled by means of a rheostat, in the primary circuit of the transformer and by insertion of a high resistance of 400,000 ohms between the transformer and the " live " side of the condenser, and purification and manipulation of the gases was in accordance with standard practice.

ACTIVE NITROGEN

1—In all previous experiments, the highest concentration of active nitrogen which has been measured chemically has not exceeded some 2%.† The author has, however, obtained up to 12% by admixture of excess (75%) of helium with the nitrogen before sparking, an effect which is presumably attributable to a collision of the second kind. It is inadvisable to attempt the formulation of a definite reaction process here, but it may be significant that the excited molecular ion N_2^+ can dissociate adiabatically to N and N^+ at 24 V, which is so close to the ionization potential of helium that the electron interchange reaction $He^+ + N_2 \rightarrow He + N + N^+$ would seem very probable. Little assistance can here be derived from the observations of Duffendack and Wolfe,* since their studies were made upon helium with a trace of nitrogen and excited by a low-voltage arc, so that the conditions of experiment are not comparable.

---

* *Cf.* Peters and Pranschke, 'Brennstoff-Chem.,' vol. 11, p. 239 (1930).

† Strutt, 'Proc. Roy. Soc.,' A, vol. 86, p. 56 (1912); Elod, 'Dissertation, Tech. Hochsch., Karlsruhe.,' 1914; Willey and Rideal, 'J. Chem. Soc.,' p. 1804 (1926).

The recombination of the atoms thus obtained being certainly a termolecular process, the third body being specifically a molecule of nitrogen, their life should be prolonged in the presence of helium, so that this gas will appear to act as a catalyst for the production of active nitrogen and an anticatalyst against its decay.

In these investigations, the active nitrogen was determined with nitric oxide,† while the afterglow was followed with a photocell.‡ The luminosity was found initially feebler but of longer duration than when nitrogen with the usual trace of photogen,§ but otherwise uncontaminated, was used; the amount of chemically active gas similarly increased above the normal value as indicated above. It may be mentioned that with 90% and over of helium the afterglow was a vivid brick-red in colour. In the above experiments condensed A.C. discharges were employed.

2—(*a*) With the impulse discharge and the ordinary photogenic gas, the concentration of active nitrogen increased in a most marked manner, up to 40% being obtained when nitric oxide was used as the reagent, and some 13% with various organic vapours such as pentane, toluene, etc. The difference in these values may be due either to some purely catalytic destruction of the active gas by the reagent, as suggested by Strutt,‖ or else to the assumption of an incorrect reaction mechanism for the first process;¶ but the latter figure admits of no uncertainty as to the minimum value, since it is calculated from the yield of hydrogen cyanide, the principal product of the interaction of active nitrogen and such hydrocarbons.** It is probably a little low, since the excess of the reagent, condensed out along with the hydrogen cyanide in a trap cooled with liquid air, had a most unpleasant odour suggestive of nitriles, even after all of the main product had been removed by a current of air, the trap being cooled with ice in the meantime.

---

* 'Phys. Rev.,' vol. 34, p. 412 (1929).

† *Cf.* Willey and Rideal, *loc. cit.*, and 'J. Chem. Soc.,' p. 283 (1927).

‡ 'J. Chem. Soc.,' p. 336 (1930).

§ This word, derived from the Greek, has been coined by the author to denote those traces of impurity—not necessarily oxygen—which must be present in nitrogen to enable it to emit its afterglow. 'J. Chem. Soc.,' p. 336 (1930).

‖ 'Proc. Roy. Soc.,' A, vol. 86, p. 58 (1911).

¶ Willey and Rideal, 'J. Chem. Soc.,' p. 669 (1927).

** Strutt, 'Proc. Roy. Soc.,' A, vol. 88, p. 544 (1913).

(*b*) When the measurements were made under fixed conditions of gas flow, pressure, distance below the discharge at which the reagent was admitted, etc., the yield of active nitrogen was independent of f (provided that f exceeded f (min)), the length of the discharge (and hence V), the magnitude of the individual current pulse (*i.e.*, $\frac{1}{2}CV^2$), and the presence of inductances of up to 10 microhenries in the discharge line, being determined only by the R.M.S. current. For instance, if a 25% yield were obtained with $f = 25$ and an R.M.S. current of 3 A, increase of f to 50, with double the length of discharge, gave the same figure provided that the condenser capacity were adjusted so that i_{RMS} remained the same as before. The yield of active gas was directly proportional to the R.M.S. current up to about 35%, but then increased less rapidly, showing signs, so far as could be judged, of becoming constant at about 50%.

(*c*) The glow intensity and electrical conductivity of the active nitrogen did not exhibit any marked increase in passing from the condensed A.C. discharges employed in other experiments to the impulse discharges used here, in confirmation of the view that these characteristics are in no way a guide as to the amounts of chemically active gas present.*

(*d*) Since Steiner has recently claimed that hydrogen reacts with active nitrogen to give small amounts of hydrazine,† attempts were made to repeat the observation, but without success, although a 24-hour run was given. This agrees with the results of other workers.‡

(*e*) A previous observation that ammonia is decomposed by active nitrogen§ has been reinvestigated at the suggestion of a collaborator, Mr. W. A. Stringfellow, since this should not occur in a pure collision of the second kind, because the heat of the N-H linkage (88 k cals) exceeds the value of about 50 k cals previously found for the energy of active nitrogen. It was found that the earlier result could not be repeated if precautions were taken to eliminate all stray discharges, and it appears that the effects noted before were spurious. Similar trouble has recently been reported by Rodebush and Wahl,‖ in studies upon the appearance of the (OH) band in sparked water vapour or hydrogen-oxygen mixtures at low pressures. The experiment was extended to include an examination of the afterglow spectrum, in the presence of ammonia; but exposures of at least 24 hours, using Ilford hypersensitive panchromatic plates (H & D 2700) with a Hilger E3 instrument, did not show any marked change as compared with the usual bands, although a greenish tinge often appeared in the luminosity under visual examination when considerable amounts of the diluent were being added, while if too much ammonia were present the glow completely vanished. The author inclines to the view that surface conditions may have some influence here.

---

* *Cf.* Willey, 'J. Chem. Soc.,' p. 2831 (1927); and other references there cited.

† 'Z. Elektrochem.,' vol. 36, p. 807 (1930).

‡ Lewis, 'J. Amer. Chem. Soc.,' vol. 51, pp. 654, 655 (1929); Strutt, 'Proc. Roy. Soc.,' A, vol. 85, p. 219 (1911); Willey and Rideal, 'J. Chem. Soc.,' p. 669 (1929).

§ Willey and Rideal, 'J. Chem. Soc.,' p. 669,(1929).

‖ 'J. Chem. Phys.,' vol. 1, p. 696 (1933).

3—*Power Yields*—It was hoped here, as in other cases which will be reported in later communications, to adjust the conditions so that the same yield of active nitrogen should be produced by each of the three types of discharge concomitantly with measurements of the power applied in each instance, but in practice this was found impossible because of the nature of the A.C. and D.C. supplies. The discharge tube was made of clear quartz, with aluminium electrodes, and was water-cooled; the reagent was added 10 cm from the point at which the gas left the discharge, and the streaming velocity through the 4 mm tubing used (100 cc/sec at 6 mm) was so great as to inhibit any measurable back-diffusion of the diluent into the spark. Determinations were made with nitric oxide, and the electrode adjacent to the reagent inlet was earthed to prevent any stray discharge through the mixture undergoing reaction. The results are given in Table I and show that (*a*) within the limits of experimental error, when the yield of active nitrogen is plotted against watts, a straight line through the origin for the impulse discharge is obtained, and (*b*) after a short initial stage the

TABLE I

Discharge	Watts	% Active nitrogen
A.C.	20	2·2
	30	2·7
	40	3·0
	50 to 150	3·2 (constant)
D.C.	80 to 180	6·7 (constant)
Impulse	5·0	3·2
	9·5	6·7
	14	9·6
	25	17·2
	35	24·0

line is parallel to the power axis for the A.C. and D.C. discharges. Moreover, while in the latter case the yields of active nitrogen at the lowest power inputs possible are respectively 3·2% at 50 watts and 6·7 at 80 watts, the same percentages may severally be obtained at 5·0 and 9·5 watts when the impulse discharge is employed. It may be mentioned that when the A.C. and D.C. discharges were in use, the afterglow was very feeble, as was also the flame at the confluence of the active gas and the reagent, in agreement with the author's earlier observation that practically non-luminous but chemically reactive nitrogen may be obtained with uncondensed discharges. Whether or not it is merely a coincidence that the yield with the A.C. discharge is almost exactly half that from the D.C. one cannot as yet be stated, since further data upon the point are not available. A similar superiority of the impulse discharge was found in other experiments in which part of the active nitrogen stream was allowed to flow through a tube fitted with windows for the end-on observation of the glow, while the remainder passed into a vessel carrying two thermometers, one with its bulb uncoated and the other covered with reduced copper (Table II). The difference in the readings of the two thermometers is a measure of the amount of active nitrogen present, while the intensity of the luminosity may be followed by means of photocells aligned upon the windows.* No data as to the relation between glow-intensity and amount of thermogenerative species can be derived from these experiments, since it was not found possible to adjust the conditions so that the luminosity completely vanished while the gas was in the observation tube; but the results are very convincing as to the yields of active gas obtained with comparable watt-inputs as we pass from the A.C. and D.C. to the impulse discharges.

TABLE II

Discharge	Watts	Glow-intensity	Temperature difference
A.C.	20	0·3	0·7
	35	0·5	1·2
D.C.	29	0·7	1·7
	56	1·0	2·4
Impulse	21·5	12·3 (Initial)	14·0 (Initial)
		7·4 (Final)	12·5 (Final)
	8·3	5·6	9·5

It will be noticed that in the first of the impulse discharge experiments the glow intensity at the beginning of the run was much more than at the end, although the heating effects were but little changed. This would appear an interesting example of the influence of the walls upon active nitrogen. The apparatus had been newly constructed for the purpose of these experiments, and, according to the usual practice of the author, was not baked out under a high vacuum but allowed to " run in " for some hours before the quantitative work commenced; this is generally much preferable to the more drastic procedure alluded to, since the active nitrogen appears able of itself to produce the desired effect if given adequate time. Apparently, however, the final steady state had not been reached when the readings started, but set in during the hour which elapsed between the first and second observations; such a sudden onset of the desired constancy of conditions has often been observed by the author in many experiments involving discharges. The phenomenon indicates, however, the invalidity of the assumption that glow-intensity is of necessity a guide to the amount of chemically active nitrogen present.

---

* Willey, 'J. Chem. Soc.,' p. 2831 (1927).

DISCUSSION

It does not seem possible, in the absence of very complete quantitative data upon the spectra developed in nitrogen, when equivalent amounts of power are applied in different ways, as in these experiments, to offer definite explanations as to the very great superiority of the impulse discharge over the other types, both as regards absolute and power yields of the active nitrogen. The following considerations may, however, be offered.

Uncondensed discharges in nitrogen exhibit mainly the first, second, and fourth positive band systems, corresponding to transitions—

first positive	$B^3\Pi \rightarrow A^3\Sigma$
second positive	$C^3\Pi \rightarrow B^3\Pi$
fourth positive	$D^3\Sigma \rightarrow B^3\Pi$

the last named being, as a rule, much less noticeable than the first two. It is thus clear that in the absence of disturbing collision effects, nitrogen molecules having radiated these bands will ultimately exist in the metastable $A^3\Sigma$ state, which, according to the latest data, lies some 6V above the ground state.[*] The older value for the A – X transition being 8·1 V, it was possible, although clearly not very probable, as suggested by Cario and Kaplan,[†] for metastable atoms in the 2P state (3·6 V) to be formed in the

[*] Cario, ' Z. Physik,' vol. 89, p. 523 (1934).

[†] ' Z. Physik, vol. 58, p. 769 (1929).

bimolecular reaction $2N_2\ A^3\Sigma \rightarrow N_2 + 2N\ (2P)$, the α-bands of the afterglow then resulting from the interaction of the metastable atom, and another metastable molecule in a collision of the second kind. The infra-red bands result from a similar excitation by the metastable 2D atom, 2·35 V. Atoms may thus be expected in a gas which has been subjected to only a mild discharge, in consonance with the observations that chemical activity is detectable under these conditions,* while if the current be sufficiently raised, the afterglow also appears.†

Condensed discharges, on the other hand, show a stronger development of the fourth positive bands, the upper states of which are unstable and permit of adiabatic dissociation of the molecule $N_2\ (D^3\Sigma) \rightarrow N\ (2D) + N\ (2P)$, while the presence of spark lines as well indicates that the excited molecular ion at 24 V is also very probably breaking up, as already indicated (p. 165). Moreover, weak discharges through active nitrogen show a marked resemblance to the condensed discharges which all authorities agree are the most effective for the production of the glowing gas.‡ It would thus appear that one reason for the effects observed in these experiments lies in the enhanced production of those unstable upper states of the $D^3\Sigma$ level of N_2 when the controlled condenser discharges are in use, but the destruction of the atoms by a second discharge observed by Wrede (*loc. cit.*) and others (Strutt, *loc. cit.*; Tiede and Chomse§) suggests that the very short duration of the impulse discharge is also very beneficial in preventing their excitation to higher levels in which they can either recombine more easily than usual, probably in purely bimolecular reactions, or cease to exhibit the characteristics which give active nitrogen its somewhat unique properties. Hence we may suppose, on the balance of evidence, that the impulse discharge favours first the initial production of atoms through inducing those states of the parent molecule in which it dissociates most easily, and later conserves them, as we have seen; whereas the other types are both much less fruitful as regards the primary yield and also highly destructive, by reason of their continuous nature, upon the small amounts of reaction product actually generated.

It will be shown, in later communications, that this time-factor is of great importance in other impulse-discharge reactions.

These experiments were carried out in the Sir William Ramsay Laboratories, University College, in London, 1931–33, and it is a pleasure to

---

* *Supra*, and 'J. Chem. Soc.,' p. 2831 (1927).

† Strutt, 'Proc. Roy. Soc.,' A, vol. 92, p. 438 (1916).

‡ Bay and Steiner, 'Z. phys. Chem.,' B, vol. 9, p. 93 (1930).

§ 'Ber. deuts. chem. Ges.,' vol. 63, p. 1839 (1930).

acknowledge the facilities extended by Professor Donnan and the College authorities through an Honorary Research Associateship. I am also indebted to Mr. W. A. Bayliss for assistance in the experimental work, and very particularly to Major F. A. Freeth, F.R.S., Dr. H. M. Barlow, and Dr. S. G. Foord for their advice. criticism, and interest.

Dipole Induction and the Solvent Effect in Dipole Moment Measurements

By F. C. Frank

(*Communicated by N. V. Sidgwick, F.R.S.—Received April* 17, 1935)

1—Introduction

It has been conclusively proved in the last two years that there is a significant effect of the solvent on the measured value of dipole moment of a solute.[*][†] It has also been shown that the effect is mainly determined by the dielectric constant of the medium, suggesting a relatively simple physical cause, but the precise law of relationship is still disputed. A list of the formulae which had been proposed at the time of the Faraday Society's discussion is collated by Goss.[‡] These all have the status of empirical formulae, since that of Raman and Krishnan contains two difficult accessible constants arising from molecular anisotropy and has not been tested with independent values of these constants. The theoretical formula of Weigle[§] is omitted from this list. It receives some discussion in the next section of this paper.

These formulae agree in disregarding the properties of the solute, excepting its dipole moment. The solvent effect is considered to be proportionately the same for all polar substances. Müller[||] makes an

[*] Müller, 'Phys. Z.,' vol. 33, p. 731 (1932) ; *ibid.*, vol. 34, p. 689 (1933).

[†] Various contributions to the Faraday Society's General Discussion on "Dipole Moments" ('Trans. Faraday Soc.,' vol. 30, pp. 677–904 (1934)), hereinafter referred to as "Dipole Moments" (1934), notably Arkel and Snoek, p. 707 ; Müller, p. 729 ; Jenkins, p. 739 ; with discussions commencing on pp. 719 and 745.

[‡] "Dipole Moments" (1934), p. 751.

[§] 'Helv. Phys. Acta,' vol. 6, p. 68 (1933).

[||] 'Phys. Z.,' vol. 34, p. 689 (1933).

exception in suggesting a modified law for certain substances, mainly ketones. The hydrogen halides, with positive solvent effects, are considered entirely anomalous, and their increase of moment is attributed to an intramolecular change towards more ionic linkage.*

It is clearly necessary to understand as fully as possible the nature of the general phenomenon before such chemical deductions as the latter may be held with confidence. Moreover, hydrocarbon radicals in the molecule are in many ways analogous to solvent molecules (*e.g.*, solutions of normal alcohols in heptane, hexane, or cyclohexane all give practically superposable curves of P_0 against concentration of hydroxyl groups per cc†). It follows that effects of solvents and substitution must be considered together.

The field around a dipole molecule is quite intense, of the order 10^7 volts/cm. Various authors have considered the effect it can have on the medium. Müller‡ suggests that it causes partial dielectric saturation in the medium. There are two objections: firstly, that it does not explain why the dielectric constant should be the principal parameter for the solvent; secondly, that this saturation is not shown in the refractivities. So that it must be entirely in the relatively trivial atom polarization if present at all in non-polar solvents. Weigle§ derived a theoretical formula which may be expressed as:

$$\Delta\mu = \mathrm{A}\,\frac{\mu^3}{a^6}\,n\,\frac{(\alpha_1 - \alpha_2)^2}{k\mathrm{T}} + \mathrm{B}\,\mu n\,\frac{(a_1 + 2\alpha_2)}{3},$$

where A and B are constants depending on the geometry of the solute molecule. *The first term is the only one considered in detail, and arises* from the orientation of anisotropic solvent molecules in the dipole field. It will not normally account for a value of $\Delta\mu/\mu$ larger than 1%. It is negative for a spherical molecule. The second term is zero for a spherical molecule, but can be much larger than the first (it is calculated as 15% in a particular case). It arises directly from the polarization of the medium by the dipole. This is the basis of the treatment that is used in the present paper, though the conclusions diverge in several respects from those of Weigle. Higasi‖ has interpreted some experimental results in terms of Weigle's theory. These refer to the change of measured dipole moment from benzene to ether as solvents, which gives a fall for chlorobenzene, acetone, nitrobenzene and propyl chloride, but a rise for various alcohols and for aniline. He explains that Weigle's theory (the second term) predicts a negative solvent effect for molecules elongated on the dipole axis and a positive solvent effect for molecules elongated perpendicular to this axis. The latter molecules should have negative Kerr constants, and this is so for all but one of the molecules showing increased moments in ether. Table I is taken from Higasi.

---

* Fairbrother, " Dipole Moments," p. 862 (1934).

† Snoek, " Dipole Moments," p. 772 (1934).

‡ 'Phys. Z.,' vol. 33, p. 731 (1932).

§ 'Helv. Phys. Acta,' vol. 6, p. 68 (1933).

‖ 'Abstracts Bull. Inst. Phys. Chem. Res.,' p. 65 (1934), (with 'Sci. Pap. Inst. Chem. Res.,' Tokyo, vol. 25, Nos. 525-28).

TABLE I—ELECTRIC MOMENTS

Solvent Solute	C_6H_{14} 1·91	C_6H_6 2·28	CS_2 2·64	Et_2O 4·35	Kerr constant
C_6H_5Cl	1·62	1·54	1·46	1·3	3·85S red
$(CH_3)_2CO$	2·81	2·76	2·70	2·2	5·05S white
$C_6H_5NO_2$	4·05	3·96	3·74	3·2	100·7S white
$n - C_3H_7Cl$	1·99	1·92	1·81	1·7	2·43S red
C_2H_5OH	—	1·74	—	1·8	0·23S yellow
iC_4H_9OH	—	1·72	—	1·8	$-$ 1·375S red
$C_5H_{11}OH$	—	1·66	—	1·8	$-$ 0·984S red
$C_6H_5NH_2$	1·50	1·54	—	1·67	$-$ 0·507S white

(S is the Kerr constant for CS_2 at the same wave-length.)

It is difficult to be sure that the influence of molecular association is absent from benzene solutions of the alcohols, nor is ether a solvent entirely free from suspicion of anomaly, but the conclusions of Higasi are essentially identical with those of the present paper, which are supported below with evidence free from these criticisms.

The theory of Raman and Krishnan,* like that primarily considered by Weigle, is also concerned with molecular anisotropy, but introduces it by a modification of the Clausius-Mosotti formula.

Quite different from these theories is that outlined by Debye,† who supposes that the local field acting on a molecule due to its neighbours can be regarded as directed along an axis which, owing to the quasi-crystalline structure of liquids, moves slowly compared with the thermal motion of liquids and independently of fields applied from without for the purpose of measurement. The local field then gives another energy term beside that of thermal agitation tending to hinder the orientation of dipole molecules by the externally applied field. It seems not improbable that this may prove a valuable theory for polar liquids, such as water, but less

* Goss, "Dipole Moments," p. 751 (1934).

† 'Phys. Z.,' vol. 36, p. 100 (1935).

likely that it will explain behaviour in non-polar solvents. It does not suggest why the dielectric constant should be the principal parameter for the solvent, nor can it explain any but uniform negative solvent effects.

2—THE PHYSICAL MODEL

When a polar molecule is placed in a polarizable medium, it sets up an induced polarization in the medium, and we shall show that in general there is a net resultant moment of induced polarization. This may or may not be along the axis of the primary dipole, and when along the same axis may be in the same or the opposite direction, but it is fixed relative to the primary dipole and independently of any motion of the medium, so that it has to be vectorially combined with the primary moment. It is part of the dipole moment of the molecule so long as the latter remains in the medium. This is the chief effect considered herein. There are several other effects arising from the action of the dipole field on the medium, *e.g.*, an orientation of anisotropic or polar molecules in the medium, and an attraction of solvent molecules to the polar molecule resulting in an increase of density in its neighbourhood and an actual binding to it of an approximately monomolecular layer of solvent. We shall be in a better position to examine these secondary effects at a later stage (*see* § 5).

The induced moment in a solvent molecule in the neighbourhood of a dipole molecule is

$$\mu' = \alpha \mathrm{F} = \alpha \mathrm{E}\,\frac{\varepsilon + 2}{3}$$

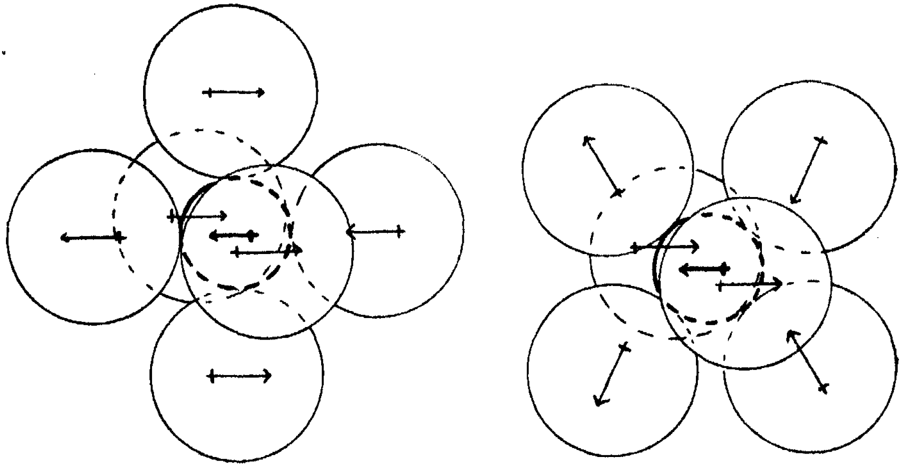
$$\mathrm{E} = c\,\frac{\mu}{\varepsilon r^3},$$

therefore

$$\mu' = \frac{c\mu\alpha(\varepsilon + 2)}{3r^3\varepsilon}.$$

(Here α is the polarizability of the solvent molecule, ε the dielectric constant of the medium, μ the moment of the polar molecule, r the distance between them, and c an orientation factor varying between 1 and 2.) If we insert values $\alpha = 10^{-23}$ cc, $\varepsilon = 2 \cdot 3$ (approximate values for benzene), $\mu =$ 1 D, $r = 3$ A, $c = 2$, we get $\mu' = 0 \cdot 46$ D. This calculation is in error through supposing the polarizability of the molecule located at a point, among other reasons, but the order of magnitude is correct.

To a considerable degree, however, the induced moments in the neighbouring solvent molecules cancel each other out, as is easily seen by reference to fig. 1, in which, for simplicity, the neighbouring solvent molecules are represented as being cubically packed. The dipole molecule is shown in the middle and induced moments are shown by thin arrows. In fig. 1*a*, for example, there are four induced moments opposed to the primary moment and two augmenting it, but each of the latter is twice as large as each of the former, so that the net effect is zero. The effect of further layers of solvent will be considered later, the medium being treated as a continuum. For the present, let us merely say that whatever its effect may be it is proportionately the same for all solute molecules. It does not influence our conclusions about the difference in solvent effect for molecules of different shapes, but may introduce a universal term which, for the present, we are neglecting.



(*a*) (*b*)

FIG. 1.

With this proviso the solvent effect arising directly from polarization of the solvent is zero for the spherical molecule. HCl and $H_2C = O$ approximate to this simplest case, but we have no measurements for the latter, while the former has other complications.

Polar molecules with large hydrocarbon radicals can be imagined as formed from the above systems by simply uniting the dipole molecule to one of its neighbouring solvent molecules, *e.g.*, we form a model of chlorobenzene dissolved in benzene by uniting the right-hand molecule to the central molecule in fig. 1*a*, or of benzophenone in benzene by uniting the two right-hand molecules of fig. 1*b* to the central molecule. This causes a certain contraction but otherwise little change in the system, except, most importantly, that the phenyl group goes with the dipole into the vapour or into another solvent, whereas the benzene molecules do not. It is the induced moment in these benezene molecules alone which we have to consider in reckoning the solvent effect. In these two molecules it is negative, *i.e.*, opposing the primary moment. Hence, a higher value will be found for the moment in the vapour state and a lower value if the benzene is replaced by some more polarizable solvent, *e.g.*, CS_2. It is important to notice that the sign of this change depends on the position of the hydrocarbon radical relative to the dipole. The nearest approximation to the opposite case, wherein radicals are fixed in the plane at right angles to the dipole axis, is found in the tertiary amines, and, accordingly, the moment of triethylamine is higher in solution than in the vapour state (*see* § 3d).

An attempt is made at a mathematical generalization of these conclusions, and an expression is derived for the induced moment in any portion of the medium around the dipole, the medium being regarded as continuous. This, with other approximations in the model, is discussed immediately after the calculation (Notes (*a*) to (*c*)). The calculation is incomplete in that it fails to decide the shape of molecule with zero solvent effect arising in this manner (*see* note (*d*)).

The potential in the neighbourhood of a dipole of moment μ and length a, centred at the origin and directed along the x axis, in a medium of dielectric constant ε, is

$$V = \frac{\mu/a\varepsilon}{\sqrt{(x-a)^2+y^2+z^2}} - \frac{\mu/a\varepsilon}{\sqrt{(x+a)^2+y^2+z^2}}.$$

When a is negligible this becomes

$$V = \frac{\mu x}{\varepsilon\,(x^2+y^2+z^2)^{3/2}} = \frac{\mu\cos\theta}{\varepsilon r^2}, \tag{1}$$

expressed in polar co-ordinates referred to the same origin and the x axis.

The component of field parallel to the x axis is

$$E_x = \frac{-\delta V}{\delta x} = \mu\,.\,\frac{3\cos^2\theta - 1}{\varepsilon r^3}. \tag{2}$$

In the plane xy the component parallel to the y axis is

$$E_y = \frac{-\delta v}{\delta y} = \mu\,.\,\frac{3\sin\theta\cos\theta}{\varepsilon r^3}. \tag{3}$$

E_x is O when

$$3\cos^2\theta = 1,$$

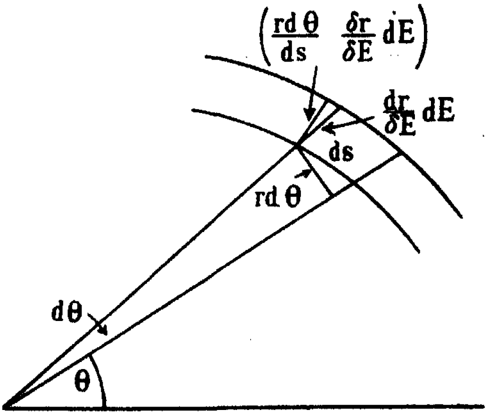
i.e., when $\cos\theta = \pm 1/\sqrt{3}$
i.e., $\theta = \pm 54{\cdot}7°$.

The total resultant intensity at any point is

$$E = \sqrt{\left(\frac{\delta v}{\delta x}\right)^2 + \left(\frac{\delta v}{\delta y}\right)^2}$$

$$= \frac{\mu}{\varepsilon r^3}\sqrt{1 + 3\cos^2\theta}. \qquad (4)$$

This equation will define a surface of uniform field strength. It is convenient for the purpose of integration to divide the space around a dipole by such surfaces into elementary shells of uniform field, and further to divide such elementary shells into elementary rings wherein this uniform field is uniformly inclined to the axis. The volume of an elementary ring defined by the limits E, E + dE, and θ, $\theta + d\theta$, is



$\left(\frac{rd\theta}{ds}\ \frac{\delta r}{\delta E}dE\right)$
$\frac{\delta r}{\delta E}dE$
ds
rd θ
dθ
θ

FIG. 2.

$$-\,2\pi r\sin\theta\,.\,ds\,.\,\frac{r\,d\theta}{ds}\,.\left(\frac{\delta r}{\delta E}\right).\,dE = -\,2\pi r^2\sin\theta\,.\,d\theta\,.\left(\frac{\delta r}{\delta E}\right).\,dE.$$

From (4) we have

$$r = (1 + 3\cos^2\theta)^{1/6}\left(\frac{\mu}{E\varepsilon}\right)^{1/3},$$

therefore

$$\frac{\delta r}{\delta E} = -\,(1 + 3\cos^2\theta)^{1/6}\,\frac{\mu^{1/3}}{3E^{4/3}\varepsilon^{1/3}}$$

$$= \frac{-\,r}{3E}.$$

Therefore the volume of the elementary ring is

$$\frac{2\pi r^3}{3E}\sin\theta\,.\,d\theta\,.\,dE. \qquad (5)$$

The induced moment per unit volume of the medium is

$$\mathrm{I} = \frac{\mathrm{E}\,(\varepsilon - 1)}{4\pi}. \tag{6}$$

For the resultant induced moment in an elementary ring the components in the y direction sum to zero, so that the resultant moment is

$$\begin{aligned}
&\frac{\mathrm{E}_x\,(\varepsilon - 1)}{4\pi} \cdot \frac{2\pi r^3}{3\mathrm{E}} \sin\theta \,.\, d\theta \,.\, d\mathrm{E} \\
&\qquad = \frac{\varepsilon - 1}{4\pi} \cdot \frac{2\pi r^3}{3\mathrm{E}} \cdot \frac{\mu}{\varepsilon r^3} \,.\, \sin\theta\,(3\cos^2\theta - 1) \,.\, d\theta \,.\, d\mathrm{E} \\
&\qquad\qquad\qquad \text{(by introduction of (2))} \\
&\qquad = \frac{(\varepsilon - 1)\,\mu}{6\mathrm{E}\varepsilon} \,.\, \sin\theta\,(3\cos^2\theta - 1) \,.\, d\theta \,.\, d\mathrm{E}. \qquad (7)
\end{aligned}$$

Integrating this with respect to θ we get

$$\frac{(\varepsilon - 1)\,\mu}{6\mathrm{E}\varepsilon} \,.\, d\mathrm{E} \,.\, {}_{\theta_0}^{\theta_1}\Big[\cos\theta - \cos^3\theta\Big]. \tag{8}$$

Over the whole shell (limits of θ, 0 to π) this reduces to zero.

This result is not unique for the particular sort of surface chosen, but is equally true for spherical surfaces, or, indeed any family of surfaces differing only by a scale-factor. It shows that the resultant induced moment in every completed shell of uniform field in the medium is zero. The only shells with a finite contribution to the moment are those rendered incomplete by the boundary conditions, which come under two heads : (i) the boundary of the polar molecule; and (ii) the external boundary of the solution. These are purely geometrical considerations without scale-factor, so that we may not dismiss (ii) on the grounds that this boundary is too far away to matter. To do so would leave us the problem of deciding whether to integrate to infinite distance or to infinitely small field strength, the former by spherical surfaces or the latter by the surfaces defined above; there is a finite difference between the moments induced in the solvent enclosed within these two surfaces.

What we require for a body of rotation is the double integral

$$\Delta\mu = \iint \frac{(\varepsilon - 1)\,\mu}{6\mathrm{E}\varepsilon} \,.\, \sin\theta\,(3\cos^2\theta - 1)\, d\theta \,.\, d\mathrm{E},$$

which must necessarily be of the form

$$\Delta\mu = \mathrm{A} \,.\, \mu \,.\, \frac{\varepsilon - 1}{\varepsilon}$$

(this is true for all bodies besides those of rotation). Here A is a numerical factor which can be divided into two parts, $A = (A_1 + A_2)$, where A_1 depends only on the geometry of the polar molecule (without reference to its size) and A_2 depends similarly on the external boundary conditions. We shall not at present enter into the problem of evaluating A_2 (*see* note (*d*)).

A_1 must be evaluated by a graphical method for any usual molecular geometry. The procedure is as follows:

Approximate Graphical Integration.

This is applicable directly to molecules which are true bodies of rotation, and with a modification to all molecules having symmetry about the dipole axis. It may be used to compute the x component of induced moment when the molecule is unsymmetrical, but further computation will be needed to obtain the y component.

The resultant moment in an elementary ring (r, θ) with cross-section a^2 is $I_x \times$ volume of ring

$$= 2\pi a^2 r \sin\theta \,.\, \frac{\varepsilon - 1}{4\pi\varepsilon} \,.\, \mu \,.\, \frac{3\cos^2\theta - 1}{r^3}$$

$$= \frac{\sin\theta\,(3\cos^2\theta - 1)}{2\,(r/a)^2} \,.\, \frac{(\varepsilon - 1)}{\varepsilon} \,.\, \mu.$$

This is the basis for the following procedure for integration:

Plot the curve of uniform field as a polar graph:

$$r^3 = K\sqrt{1 + 3\cos^2\theta}$$

(K being any convenient factor).

Divide the area within this curve into unit elements of area. Ascertain the co-ordinates (r, θ) for the mid-point of each element of area. r must be measured in units of length corresponding to the unit elements of area. Compute

$$\frac{\sin\theta\,(3\cos^2\theta - 1)}{r^2}$$

for each element, and label the element with this value. It is convenient to draw a graph of the trigonometrical factor first. The diagram now prepared can be used for computing A for any molecule.

Draw the axial section of the molecule (a body of rotation, or approximated as one) on the diagram, with the dipole centre at the origin and the dipole directed along the axis, to any scale, so that it lies wholly within the boundary curve of the diagram. It should be as large as possible for greatest accuracy. The sum of the values $\sin\theta\,(3\cos^2\theta - 1)/r^2$ for all the elements of area lying between the boundary of the molecule and the boundary curve of the diagram, on one side of the axis, is equal to $2A_1$, where A_1 is the constant dependent on the molecule in the equation

$$\Delta\mu = (A_1 + A_2)\,\mu\,.\,\frac{\varepsilon - 1}{\varepsilon}\,.$$

Notes on Approximations

(*a*) The finite length of the dipole is neglected. Its actual length is difficult to estimate ; indeed, it is doubtful whether it has any precise meaning, since probably the field can only be truly regarded as a dipole field at distances great enough to make the dipole length practically negligible. But we may obtain an order of magnitude by supposing that a moment of 1·0 D arises by displacement of two electrons forming a valency link. This displacement is then 0·105 A. A reasonable finite length for the dipole will not grossly alter the results of the calculation.

(*b*) The medium is regarded as a uniform continuum—There is a precedent for this in the treatment of the medium between ions as continuous in the Debye-Hückel-Onsager theory of electrolytic conduction. If this may be done successfully even in so complicated a medium as water, it is not so very unjustifiable in the hydrocarbon solvents.

(*c*) The field distribution is considered to be always the same as it is in a uniform medium—This implies that the body of the polar molecule is a uniform medium with the same dielectric constant as the solvent, which cannot be even approximately true in all solvents. It is also untrue in the gaseous state. For the sort of molecule we have mainly to deal with, the bulk of which is a hydrocarbon radical, this distortion of the field is least in the common non-polar solvents with dielectric constants around 2–2·5.

Qualitatively the chief result of this error is that molecules considered to have zero solvent effect according to the simple deduction will actually show a positive effect. This may be seen by considering the influence of the reflex polarization in the polar molecule induced by induced moments in its neighbours, which is always in the same direction as the primary moment (consider fig. 1).

Under the same head we must consider the effect of anisotropic structure in the polar molecule. If it contains an aliphatic chain of several members the induced moment in the portion of the chain remote from the dipole is mainly the result of secondary induction from induced moments, *i.e.*, an effect relayed along the chain, especially when the molecule is in the gas. Consequently, its direction is by no means simply related to the direction of the primary moment, and is, in fact, very sensitive to slight changes in the configuration of the chain. In the gas phase, therefore, there may be a rather erratic change of moment with chain length in a homologous series. This will be less evident in solution where other molecules besides the atoms of the chain take part in relaying the polarization, so that we shall have some degree of anomaly in the solvent effect. There is some evidence that this occurs in the series of normal aliphatic ethers, *see* § 4.

(*d*) The external boundary conditions—We cannot immediately dismiss even the possibility that A_2 depends on the shape of the condenser used for measurement but, fortunately, the requirements of dielectric constant measurement enforce general similarity of form in these condensers. The experimental values indicate that A_2 is small but do not certainly discriminate against the spherical surface which Weigle regards as that outside which there is no resultant induced moment. The surface of uniform field strength remains in any case a convenient one for calculation, as it is the appropriate surface for changes in the secondary effects of the field, *e.g.*, compression of the solvent, with consequent rise of dielectric constant, by dipole attraction, *see* § 5 (*a*) and (*b*).

3—EXAMPLES OF SOLVENT EFFECT CLASSIFIED ACCORDING TO MOLECULAR GEOMETRY

a—*Small Molecules with no Large Group Attached, e.g., HCl, H_2O*

Hydrogen halides—The data given in Table II are as collated by Fairbrother,* who has made all of the measurements in solution.

These figures leave no doubt that there is an increase of the moment of these substances of the order 10% or more, on going from gas to solution in benzene or CCl_4. For HCl the increase is 19% if Van Vleck's value is used, or 7% with Braunmühl's measurement. The latter is probably too low. The other hydrogen halides give about the same changes of moment ($\Delta\mu$) and therefore larger proportional changes ($\Delta\mu/\mu_0$).

The graphical integration for a spherical molecule gives $A_1 \simeq +0{\cdot}09$, corresponding (if A_2 is neglected) to $\frac{\Delta\mu}{\mu} \simeq +5\%$ in benzene. This is not much dependent on the location of the dipole within the sphere provided that it is directed along a diameter. Of course, μ here is the primary moment which may be very different from μ_g, the moment in the gas. $\Delta\mu$ calc. will be reduced if we give the sphere a protuberance for the hydrogen envelope.

---

* "Dipole Moments," p. 867 (1934).

TABLE II

Molecule.	Moment	Method	Worker
HCl	1·03 D	Gas, T.	Zahn.
HCl	1·06	Gas, Opt.	Van Vleck, Zahn.
HCl	1·18	Gas	Braunmühl.
HCl	1·26	C_6H_6, Opt.	Fairbrother.
HBr	0·79	Gas, T.	Zahn.
HBr	0·80	Gas, Opt.	Van Vleck, Zahn.
HBr	1·01	C_6H_6, Opt.	Fairbrother.
HBr	0·96	CCl_4, Opt.	Fairbrother.
HI	0·38	Gas, T.	Zahn.
HI	0·41	Gas, Opt.	Van Vleck, Zahn.
HI	0·58	C_6H_6, Opt.	Fairbrother.
HI	0·50	CCl_4, Opt.	Fairbrother.

Thus we can explain the result (while neglecting A_2) by regarding the moment of HCl in the gas as a large primary moment offset by a large induced moment within the molecule. This model of the HCl molecule corresponds closely to that of Debye,* namely, a chlorine ion with a proton embedded in it, making a primary moment of $4{\cdot}77 \times 10^{10} \times 1{\cdot}27 \times 10^{-8}$ e.s.u. = 6·06 D, which is offset by an opposing induced moment in the chlorine ion.

The main source of the solvent effect in the hydrogen halides, however, is probably that described in § 2, note (*c*), as the reflex polarization of induced moments in the solvent upon the central molecule. It is in agreement with this that the proportional effect increases with increasing polarizability in the series. It appears unnecessary to have recourse to the hypothesis of Fairbrother† or Bernal,‡ namely, a partial change towards ionic linkage or a small displacement of the proton.

Water—Bernal, on the basis of his theory mentioned above, predicted that water should behave like the halogen hydrides in showing an increased moment in solution, and to a larger degree. Müller replied that in fact it did not, but he relied on Williams's earlier measurement of the moment of water dissolved in benzene, 1·7 D. His more recent value is 1·9 D, also found by himself and by Linton in dioxane. The value in ether is disputed, Hassell finding 1·97 D and Linton 1·71 D. Sänger's gas value is 1·842 ±

---

* "Polar Molecules," p. 59 (1929).

† "Dipole Moments," p. 867 (1934).

‡ "Dipole Moments," p. 872 (1934).

0·008 D. Thus water gives, if anything, a 3% increase of moment when dissolved in benzene.

The calculated increase should not be very different from that calculated for the simple model of HCl, though the tripole structure introduces complications. Fowler and Bernal* have already shown the usefulness of a spherical model of the water molecule. The difference in behaviour of the tripole and its resultant dipole is not very great for the present phenomenon. The lower polarizability of the water molecule compared with HCl tends to eliminate a source of high positive $\Delta\mu$ present in the latter.

b—*Molecules with a Radical on the Dipole Axis, e.g.*, CH_3Cl, $C_6H_5NO_2$

This class embraces practically all the substances considered by Müller in deriving his rule for the solvent effect, § 5. According to the rule, all substances should show a value of 6·7% (± 0·5%) for $\Delta\mu/\mu$ in benzene at 25°.

Methyl Chloride—The value of A_1 found by graphical integration corresponds to a value of $\Delta\mu/\mu$ in a medium of dielectric constant 2·3 of from −3 to −5%, according to the location of the dipole ($\frac{1}{8}$ of internuclear separation from Cl suggested by Meyer† or at point of contact of the combining spheres of C and Cl). There is no measured value in benzene, but Morgan's value in CCl_4 is 1·65 or 1·67 calculated from his measurement at 20° by the optical method, while the best gas value is Sänger's 1·861 ± 0·008. This makes $\Delta\mu/\mu$ − 10%. But there is some anomaly here, for practically the same value is found in hexane. In this substance, unlike HCl, the reflex polarization acts in opposite ways on two parts of the molecule.

Chlorobenzene—Calculated value of A_1 gives $\Delta\mu/\mu$ about −10% in benzene. The found value, based on 1·56 D in benzene (agreement of numerous workers) and 1·70 in the gas,‡ is − 8·2%.

Nitrobenzene—This should be very similar to chlorobenzene. The found value of $\Delta\mu/\mu$ in benzene is − 6·9% (Benzene, 3·94 D, Jenkins; Gas, 4·23 D, Groves and Sugden§).

c—*Molecules with a Single Radical not on the Dipole Axis, e.g.*, CH_3OH

The effect here depends largely on the polar co-ordinate θ of the radical. If this is less than 55° it displaces solvent with a positive induced moment, and therefore makes $\Delta\mu$ negative, as in the former case. When θ is greater than 55° the E_x component is in the opposite direction, so that this tends to make $\Delta\mu$ positive. On the other hand, there is the E_y component which reaches a maximum when θ is 55°, and raising the moment in the gas, tends to make $\Delta\mu$ negative. Thus, unless θ is fairly close to 90° a single radical cannot cause a large positive $\Delta\mu$ and the net effect will commonly be quite small.

---

* 'J. Chem. Phys.,' vol. 1, p. 515 (1933).

† 'Z. Phys. Chem.,' B, vol. 38, p. 27 (1930).

‡ McAlpine, Smyth, "Dipole Moments," p. 754 (1934).

§ 'J. Chem. Soc.,' p. 1094 (1934).

Methyl Alcohol—The best value is by Miles, 1·68 D. Wolf's value in benzene is 1·68 ± 0·02 D. Thus, $\Delta\mu/\mu$ is indistinguishable from zero. Ethyl and propyl alcohols are similar.

d—*Molecules with Radicals off the Dipole Axis, but with Axial Symmetry, e.g.*, CH_3OCH_3, $(CH_3)_3N$, $(CH_3)_2CO$

Here there is no unbalanced *y* component, so that the solvent correction is positive or negative according as θ is greater or less than 55°. If the intervalency angle in dimethyl ether is 110°, the solvent correction should be small and positive, its magnitude increasing with the distance of the dipole from the centre of the oxygen; except that the nearer the component dipoles are to the methyl groups the less justly can they be treated as one. If the valency angle is larger than 110°, the solvent correction should be significantly positive. In acetone it should be negative. In trimethylamine on the tetrahedral model θ is 70°, reckoned from the nitrogen nucleus, so that the correction must be significantly positive.

Measurements are not actually available for the methyl compounds. Ethyl compounds should be essentially similar, though complicated by flexibility.

Diethyl Ether—1·22 D is found in benzene when no allowance is made for P_A. This reduces to 1·15 D on allowing 3 cc for the latter. Meyer also found 1·15 ± 0·02 D by the solution temperature method, which is trustworthy if the solvent correction is very small.* Sänger's gas value is 1·146 ± 0·012 D. Thus the solvent correction has a value of zero or a small positive value. The measurements in this case are almost beyond dispute, numerous workers having concurred on the measurements in solution, and the gas value being the result of particularly careful work and agreeing reasonably with other less accurate measurements. There is no good reason to suppose chemical anomalies are present, so that we have here very strong evidence of the sort of dependence of solvent effect on the solute that the theory predicts.

---

* *Cf.* Jenkins, "Dipole Moments," p. 739 (1934).

Triethylamine—The published measurements certainly indicate a distinct positive solvent effect, but this depends on single measurements which are rather less certain than we should really demand to decide this important point. Diethylamine also has a positive solvent effect as we should expect; but, contrary to expectation, it appears to be larger than that for triethylamine. The gas values are supported by their similarity to the values for the corresponding methyl compounds, for which there are confirmatory measurements, *see* Table III.

Acetone—As anticipated, $\Delta\mu/\mu$ is negative. Zahn's value in the gas and Wolf's in benzene make $\Delta\mu/\mu$ about -3%. Acetone, benzophenone, and *p*-xylylene dichloride were given a separate curve of smaller slope by Müller in the original paper wherein he derived a general rule for the solvent effect, *see* § 4.

e—*Recapitulation*

The measurements which indicate solvent effects in benzene which are zero or positive are given in Table III, together with measurements for the methylamines in the gas.

TABLE III—ELECTRIC MOMENTS

Molecule	Gas	Benzene solution
CH_3OH	1·68*	1·66†
C_2H_5OH	1·70*	1·71†
C_3H_7OH	1·66*	1·66†
$(C_2H_5)_2O$	1·146 ± 0·012‡	1·15§
$(C_2H_5)_2NH$	0·94‖, 0·90¶	1·20**
$(C_2H_5)_3N$	0·76‖, 0·82¶	0·90**
$(CH_3)_2NH$	1·05‖, 0·96††, 0·90¶	—
$(CH_3)_3N$	0·60††, 0·82¶	—

* Miles; † Wolf; ‡ Sänger; § Lange, Williams, Rolinski, Meyer; ‖ Höjendahl; ¶ Ghosh; ** Williams; †† Steiger.

4—THE EFFECT OF INTRODUCING RADICALS

A hydrocarbon radical is very similar to the usual solvents for dipole measurements (alkyl to hexane, or phenyl to benzene), with the following differences:

(*a*) It is permanently anchored in position relative to the dipole.

(*b*) It is therefore effectively rather denser than average solvent of the same nature (a phenyl group is as dense as a benzene molecule, but denser than a cubic centimetre of benzene).

(*c*) It is closer to the dipole than are solvent molecules by about an Angstrom unit.

(*d*) Its formation from a solvent molecule involves the destruction of the links X—H and H—C, and the formation of a link X—C, the primary moment of which may be different from the sum of the moments of the two former links.

(*b*) and (*c*) indicate that the radical will have an induced moment rather larger than that of the solvent it displaces. The effects of induction in the radical should therefore be rather larger in magnitude and opposite in sign to the solvent effects. They should be smaller but still evident when we compare moments measured in a solvent of not too large dielectric constant. Unless effect (*d*) is large, the same classification of molecular geometry that has been made for solvent effect should apply for the influence of introduced radicals. Table IV shows this.

TABLE IV—ELECTRIC MOMENTS IN THE GAS

Molecule μ, Debye units	H_2O 1·84	CH_3OH 1·68	$(CH_3)_2O$ 1·32	
Molecule μ		CH_3CHO 2·68	$(CH_3)_2CO$ 2·85	
Molecule μ		C_2H_5CHO 2·4	$(C_2H_5)_2CO$ 2·72	
Molecule μ	HCl 1·03	CH_3Cl 1·97		
Molecule μ	NH_3 1·48	CH_3NH_2 1·23	$(CH_3)_2NH$ 1·0	$(CH_3)_3N$ 0·6

From the former discussion we should predict that, so far as the inductive effects are concerned, a radical introduced on or near the axis of the dipole should increase the moment, and radicals symmetrically introduced otherwise than on the axis should decrease or increase it according as they made an angle θ greater or less than 55°. Radicals unsymmetrically introduced require a larger value of θ to cause a decrease of moment. In agreement with this, the aldehydes and halides increase in moment with introduction of radicals, while the oxides and amines decrease. The fall seems rather too rapid in the oxides to be entirely due to the inductive effect, since we concluded from the solvent effect that θ was close to 55°. The discrepancy may be due to effect (*d*), *i.e.*, a change of primary moment or to distortion of the field, but the results definitely indicate that the primary moments of

C—X and H—X are not very different, most of the change of moment caused by introduction of a radical being due to the induced moment in the radical. The fact that the usual " bond moments " (as given, for example, by Sidgwick*) include a substantial allowance for induction in the molecule in no way invalidates their intelligent use.

The conclusions can be summarized in the rule :

The more positive the radical effect, the more negative the solvent effect, and vice versa.

Long Chain Radicals in General

It is important to notice that we only have measurements of moments in homologous series made in solution. Inductive effects are therefore only seen as second-order effects due to the difference between radical and solvent. The inductive effect in the chain can be divided into three parts:

(*a*) that in the chain close to the dipole. This is present practically unaltered from the first member of the homologous series onwards.

(*b*) that part of the chain for which θ is relatively fixed.

(*c*) the remoter parts of the chain.

Both in (*b*) and (*c*) the moment will be a little higher than that in neighbouring solvent molecules, because the induction effect is relayed to the outer parts of the radical through a path with no larger interatomic gap than 1½ A, whereas the solvent molecules receive it through at least one gap of 2½ A. Furthermore, if the chain is flexible, the moment (*c*) can have almost any orientation relative to the primary dipole.

This explains the smallness of the changes of moment in homologous series, and removes the hitherto puzzling discrepancy that measurements of dipole moment have failed to demonstrate the inductive effects through fairly considerable lengths of chain demanded both by theoretical calculation and theories of organic reaction, dissociation constants of acids, etc.† Even so, if one compares the most reliable measurements of moment all in one solvent, there is evidence of a significant change of moment with length of chain at least up to four members, *e.g.*, Partington's measurements in benzene:

Molecule	CH_3NO_2	$C_2H_5NO_2$	$C_4H_7NO_2$
μ	3·04	3·19	3·29

The same is true also of the halides and nitriles, but the alcohols and ketones show a relatively high degree of constancy, which is not easily accounted for.

---

* " The Covalent Link in Chemistry," p. 153 (1933).

† Smyth and Rogers, 'J. Amer. Chem. Soc.,' vol. 52, p. 2237 (1930); Ingold, 'Chem. Rev.,' vol. 15, p. 225 (1934).

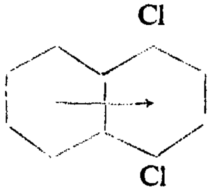
The Alkyl Ethers—The following are the moments of the alkyl ethers as measured by Sänger in the gas by the temperature method:

$(CH_3)_2O$	$(C_2H_5)_2O$	$(C_3H_7)_2O$
$1 \cdot 316 \pm 0 \cdot 012$	$1 \cdot 146 \pm 0 \cdot 012$	$0 \cdot 86$

The last value is from Sänger's earlier work in which a value of 1·10 was found for $(C_2H_5)_2O$. If the whole change is attributed to increase of the valency angle from 109·5° in dimethyl ether, it is 136° in dipropyl ether.*

But a part at least of the change must be due to induction in the radicals. This is partially cancelled in solution, and accordingly we find the change much smaller when we compare measurements in solution. We have dipropyl ether : $1 \cdot 16 \pm 0 \cdot 05$ D (Hexane, T., Meyer) and diethyl ether: $1 \cdot 15 \pm 0 \cdot 02$ (C_2H_6, T., Meyer). As we saw above, $\Delta\mu/\mu$ should not be very large for dimethyl ether. There is clearly a need for further measurements, but on those we have we can say that it is likely that the change of moment in the series is in a substantial measure due to inductive effects, and may perhaps be all due to this rather than to a change of intervalency angle.

1, 4, *Dichloronaphthalene*—This molecule has a finite moment of 0·5 D. The authors† remark that a polarization moment is not necessarily along the direction of the polar link and assume that there may be a small component of polarization moment in the direction



But Van Arkel's explanation was that any side chain, aromatic or aliphatic, on a benzene nucleus gives rise to a moment of about 0·5 D. For some reason, according to him, the unsubstituted ring in this molecule behaves as two side chains. Moore considered that the views of Mills and Nixon, together with the belief that the common atoms of the two rings in naphthalene are united by a double bond, should explain the moment.

But from the point of view of the present theory, though we cannot be certain about the electromeric effect, there must inevitably be an inductive effect in the direction suggested by Weissberger, due to the *y*-component of field in this unsymmetrical molecule. For the whole of the unsubstituted ring is in a fairly uniform resultant field in this direction. If the moment of the substance can be measured in the vapour state a considerably larger value should be found.

---

* "The Covalent Link in Chemistry," p. 148 (1933).

† Weissberger, Sängewald, and Hampson, "Dipole Moments," pp. 884, 892 (1934).

Stereochemical Calculations

It is already agreed* that for the purpose of stereochemical calculations from dipole moments one must use compounds wherein the principal dipoles are not less than 4 or 5 A away from each other. This will generally be sufficient to prevent serious interference in their solvent effects, so that stereochemical calculations remain valid. It must be understood that the value of the group moment assigned to a group X is the algebraic sum of the difference in moment of $\equiv$ C—X and $\equiv$ C—H, an inductive effect on the radical, and the solvent effect. It is thus only approximately the same as the moment of $\equiv$ C—X.

5—THE SECONDARY EFFECTS OF DIPOLE INDUCTION

a—*Attraction*

The first and most general of the secondary effects is that the polar molecule attracts to itself solvent molecules as a magnet attracts iron.

The induced moment in a solvent molecule is αF, and the resulting lowering of potential is $V_a = \alpha F^2 = \alpha E^2 \frac{(\varepsilon + 2)^2}{9}$.

$$= \frac{c^2 \alpha \mu^2 (\varepsilon + 2)^2}{9 \varepsilon^2 r^6} . \qquad (see \ \S\ 2)$$

With $c = 2$, $\alpha = 10^{-23}$ cc, $\mu = 1$ D, $r = 3$ A, and $\varepsilon = 2 \cdot 3$, this comes to $2 \cdot 13 \times 10^{-14}$ ergs $= 0 \cdot 52\, k$T at 25° C.

It follows that an approximately mono-molecular layer of solvent is practically bound to the polar molecule (though continually rearranging itself), forming a "polarization complex."

It would be surprising if the complexes of polar and non-polar molecules adhering by the attraction of induced dipoles had not given other evidence of their formation. But complex formation has already been shown for hydrogen chloride, both with krypton and with propane.† The well-known initial hygroscopy of benzene‡ is probably ascribable to the same cause.

---

* Smyth, 'J. Amer. Chem. Soc.,' vol. 54, p. 1854 (1932).

† Glockler, Roe, and Fuller, 'J. Chem. Phys.,' vol. 1, pp. 703, 709 (1933).

‡ Sidgwick, 'J. Chem. Soc.,' vol. 117, p. 1340 (1920).

We can also estimate the compression of solvent in the neighbourhood of a polar molecule due to dipole attraction.

$$\mathrm{V}_\alpha = \frac{c^2\alpha\mu^2(\varepsilon + 2)^2}{9\,\varepsilon^2 r^6}.$$

therefore

$$\frac{\delta \mathrm{V}\alpha}{\delta r} = \frac{-\,2c^2\alpha\mu^2(\varepsilon + 2)^2}{3\varepsilon^2 r^7} = \frac{-\,6\mathrm{V}\alpha}{r}.$$

Putting in the same molecular data as above, we have: attractive force per molecule

$$= \frac{\delta \mathrm{V}\alpha}{\delta r} = 4\cdot 16 \times 10^{-6} \text{ dynes}.$$

As this force falls off inversely as the seventh power of the distance, we can neglect the force on more distant molecules and, taking the molecular cross-section as 10 sq A, conclude that the pressure is about 4×10^9 dynes/sq cm = 4 kilobars. The initial compressibility $\frac{1}{v} \cdot \frac{\delta v}{\delta p}$ of benzene is $90\cdot 8 \times 10^{-6}$ per bar, but, owing to non-linear compression, the change of volume at this pressure is probably only about 10%. The corresponding change of dielectric constant is from 2·3 to 2·5. This effect operates in a similar way to the effect of anisotropy (*see* below).

b—*Anisotropic Solvents*

These have been considered by Weigle,* who concluded that the anisotropy of polarizability of solvent molecules can be directly responsible for a $\Delta\mu/\mu$ of the order 1%. The anisotropy can also influence the present effect, for in the very intense field close to the dipole the solvent is partially oriented in such a way that its most polarizable axis lies in the direction of the field. Its effective dielectric constant is therefore larger in the innermost shells than the value measured with fields of ordinary strength. Carbon disulphide (and perhaps benzene, though to a much smaller extent because it has two major axes of polarizability) may therefore behave as though its dielectric constant were larger than the known value. The solvent effect also becomes more dependent on the shape of the boundary of the molecule where it is closest to the dipole. Weigle has, in fact, considered a special case of this.

c—*Polar Solvents*

Jenkins† has claimed that when the solvent effect is allowed for according to his rule, any solvent may be used for dipole moment measurements.

* 'Helv. Phys. Acta,' vol. 6, p. 68 (1933).

† "Dipole Moments," p. 739 (1934).

The evidence for this extension to polar solvents comes from his measurement of nitrobenzene in chloroform. On the other hand, the solvent effect for HCl, which is positive in non-polar solvents, is practically zero in ethyl bromide and negative in ethylene bromide.

There appears to be no reason for excluding orientation polarization in the solvent for the present theory. But this is more easily " saturated " than distortion polarization, and saturation effects will appear in the immediate neighbourhood of the dipole. (The accurate additivity of molecular refractions, *e.g.*, of mixtures of alcohols with hydrocarbon solvents,* proves the absence of saturation effects in the distortion polarization.) It follows that the binding of solvent molecules in the polarization complex involves a loss of polarization in the solvent, which makes the solvent effect more negative. A positive solvent effect remains possible, even in a highly polar solvent, but will be rare. Polar and non-polar solvents will not behave similarly, even though their dielectric constants are equal. Jenkins's conclusion is false.

d—*Linked Ions* (*Zwitter-ionen*)

Kuhn has discussed the dielectric constant of aqueous solutions of the amino-acids† in its bearing on the problem of the configuration of the chain in such molecules. This evidence was brought forward by Wyman and McMeekin.‡ The general observation is that only substances held by chemists to have a " Zwitter-ion " structure give positive values of $d\varepsilon/dc$ (c being the concentration) when dissolved in water. In solutions of amino-acids $^{+}NH_4(CH_2)_nCO_2^{-}$ $d\varepsilon/dc$ increases linearly with n. The original deduction was that the chain was usually in its straightest position (a regular zig-zag). Kuhn, however, observing that while any deductions in so complicated a solvent are doubtful, deductions, if made, should be according to the existing equations, calculated from the experimental data a moment of 3·4 D for hexaglycylglycine, whereas the straightest form demands the much larger value 180 D. But he made no allowance at all for solvent effect. If this be made by Müller's equation, it reduces any moment to an imaginary value in a solvent with a higher dielectric constant than 4·78. Of course, Müller would not claim even an approximate extrapolation of his formula to water; but it is clear that the existing formulae favour the highly polar configuration, if any, but in truth give no decision at all.

---

* Smyth, 'J. Amer. Chem. Soc.,' vol. 51, p. 1743 (1929).

† Kuhn and Martin, 'Ber. deuts. chem. Ges.,' vol. 67, p. 1526 (1934).

‡ 'J. Amer. Chem. Soc.,' vol. 55, p. 908 (1933). These two papers give copious references to earlier work of Devoto and others.

On the basis of the present theory, we can explain the observations qualitatively with either configuration. If the chain is " straight," the increase of dielectric constant is due to the high moment of the solute offset by a negative solvent effect, mainly due to dielectric saturation. If the chain is bent round so as to bring its ends close together, the primary moment of the solute is nearly constant, and the rise of dielectric constant with *n* is due to the increase in volume of the hydrocarbon portion, which, lying to one side of the dipole, causes a positive solvent effect. The matter is complicated if the molecule is not rigid, by what Kuhn calls molecular polarization, a change of dipole length in the field. The configuration, therefore, remains undecided by this evidence.

e—*Relaxation Time for the Dipole Atmosphere*

The induced dipole in the solvent, when due to distortion polarization only, follows the motion of the primary dipole with a relaxation time which is entirely negligible. But in polar or anisotropic solvents, where the induction involves molecular orientation, there will be a relaxation time of the same order of magnitude as that for the orientation of the primary dipole. Moreover, the secondary orientation is initiated by the primary orientation and lags behind it. The system, therefore, takes a longer time to reach polarization equilibrium than the consideration of primary orientation alone predicts, and at a certain time after application of the field there exists the condition: primary orientation nearly complete, secondary orientation incomplete. When the secondary polarization is negative (*i.e.*, diminishing the primary moment), the polarization at this stage will be higher than the ultimate value. It follows that at a certain frequency just below the principal dispersion range the dielectric constant will pass through a maximum.

This is the observed form of curve for the alcohols, and for glycerine above 30° C.* It must be confessed that the secondary polarization is peculiarly complex in these hydroxylic substances.

This behaviour of the " dipole atmosphere " at high frequency is analogous to the high-frequency behaviour of ionic atmospheres, which gives rise to increased electrolytic conductivity. In a polar liquid with no applied field, the dipoles tend statistically to a special orientation one to another. When a field is applied, the molecules are oriented by the field, and the first state arrived at is not one of minimum potential energy for their mutual orientation. If the field is reversed soon enough, they never reach this state of equilibrium. The secondary orientation may give either a rise or fall in total polarization. In either case it will cause a secondary maximum on the dielectric loss curve.

---

* Ziegler, ' Phys. Z.,' vol. 35, pp. 487, 489 (1934). Mizushima, ' Sci. Pap. Inst. Phys. Chem. Res. Tokyo,' vol. 9, p. 209 (1928) (the small diagrams are inaccurate, *see* tabulated data).

6—THE LAW OF THE SOLVENT EFFECT IN ITS DEPENDENCE ON DIELECTRIC CONSTANT

Of the several laws which have been suggested, that of Müller* has the advantage of priority; this is

$$\frac{P_{0},\text{ solvent}}{P_{0},\text{ gas}} = 1 - k(\varepsilon - 1)^2,$$

where $k = 0\cdot075 \pm 0\cdot005$. This relates the orientation polarizations at infinite dilution in gas and solution to the dielectric constant of the solvent.

An examination of the material on which this is based shows that practically all the substances can be classified under § 3 b, molecules having a single radical on or near the dipole axis. Methylene chloride is an exception, but need not upset the conclusion. Measurements for acetone, benzophenone and *p*-xylylene-dichloride formed part of the original material, but were relegated to a separate $P_2 - \varepsilon$ graph showing less slope and curvature. These substances were regarded as anomalous, owing to chemical reaction with the solvent; but there is no real evidence for this, save for the ketones in carbon tetrachloride.†

On the basis of the present theory, one would rather say that Müller's rule is the form assumed by the law for molecules of a particular general shape, § 3 b, when due allowance is made for distortion of the field, strong anisotropy in one of the solvents (CS_2), etc. Even so, caution is necessary, since the other empirical equations in the field require even an opposite curvature of the $P_2 - \varepsilon$ graph. Jenkins's rule makes P_0 linear with $1/\varepsilon$, Sugden makes it linear with $1/(\varepsilon + 2)$, while Müller makes it linear with $(\varepsilon - 1)^2$. The present theory in its simplest form makes $\sqrt{P_0}$ linear with $1/\varepsilon$.

The relation between the latter rule and Jenkins's rule,

$$P_0 = C + D/\varepsilon,$$

---

* 'Phys. Z.,' vol. 34, p. 704 (1933), and "Dipole Moments," p. 732 (1934).

† [*Note added in proof, September* 18, 1935—Since the above was written, there has appeared a new paper (Müller and Mortier, 'Phys. Z.,' vol 36, p. 373 (1935)) which recognizes several distinct laws of solvent effect for different groups of substances which fall into the classification of this paper, § 3.]

is seen as follows: writing B for $A\mu$ in the simple law, § 2, we obtain from it

$$\mu_s = \mu_g + \Delta\mu = \mu_g + B\frac{\varepsilon - 1}{\varepsilon},$$

(μ_s and μ_g are moments in solution and gas respectively). Then

$$P_0 \propto \mu_s^2 = (\mu_g + B)^2 - \frac{2B}{\varepsilon}(\mu_g + B) + \frac{B^2}{\varepsilon^2}.$$

The coefficient of $1/\varepsilon^2$ will commonly be of the order of 1/10 of the coefficient of $1/\varepsilon$. If the former is neglected, we have Jenkins's rule. Jenkins* brought the respectively anisotropic and polar solvents carbon bisulphide and chloroform into line with cyclohexane and hexane at the cost of marked deviations in benzene, carbon tetrachloride, and decaline, extrapolating to an excessive moment in the gas of 5·07 D, so there is no point in studying the connexion between the two laws quantitatively.

Van Arkel and Snoek's rule was developed to deal with the change of polarization with concentration and does not recognize a dependence of moment upon solvent. Its physical basis is essentially the same as Debye's recent theory, § 1. Sugden's somewhat similar formula is entirely empirical and deals with changes both in solvent and concentration. Its validity is examined by Jenkins and Sutton.† A theoretical consideration of the complex system of solutions of finite concentration cannot be undertaken now.

The law,

$$\Delta\mu = A\mu\frac{\varepsilon - 1}{\varepsilon},$$

is tested in fig. 3 by plotting apparent values of μ, taken mainly from the work of Müller and Jenkins, against $1/\varepsilon$. Sufficient data only exist for substances of negative solvent effect. In each case a fair straight line can be drawn through values in non-polar solvents and in the gas, but the best straight line through values for non-polar solvents alone would extrapolate to a slightly higher value in the gas. This may be attributed to the effects described as distortion of the field (*see* § 2, note (*c*)), attraction of the solvent, § 5 a, etc. The substances of lower moment could equally well be fitted by smooth curves, but the points for nitrobenzene discriminate against this, showing a linear law for non-polar solvents and marked deviations in polar solvents and the highly anisotropic solvent CS_2. The value of μ to be taken for the above law is not the gas moment, but more nearly the value in a medium of dielectric constant, say, 2·5. The values noted for A are calculated with this assumption. They conform to our classification of molecular geometry and support the view that the theory given in the present paper provides the chief term in the solvent effect in infinitely dilute solutions in non-polar solvents.

---

* 'J. Chem. Soc.,' p. 483 (1934).

† (*In the press.*)

In conclusion, the author wishes to thank Dr. L. E. Sutton, Mr. W. Jackson, Mr. E. B. Moullin, Professor P. Debye, and Dr. N. V. Sidgwick for discussions and advice. He is in receipt of a maintenance grant from the Department of Scientific and Industrial Research.

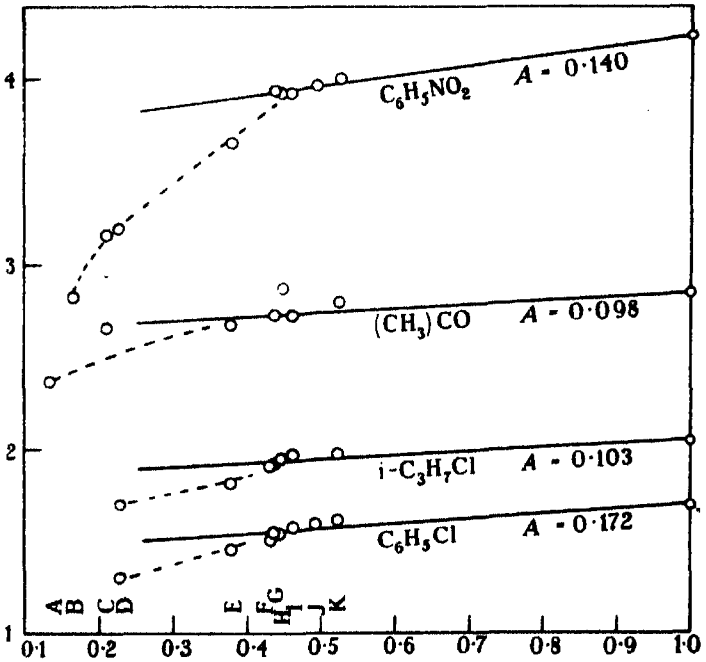


FIG. 3—A, aniline ; B, chlorobenzene ; C, chloroform ; D, ether ; E, carbon disulphide ; F, tetrachloroethylene ; G, benzene ; H, carbon tetrachloride; I, decaline ; J, cyclohexane ; K, hexane.

SUMMARY

A theory is developed accounting for the influence of the solvent upon the apparent dipole moment observed by Müller, Jenkins, and others. It is an extension of the ideas of Weigle and Higasi. It leads to a classification of molecular geometry with very different solvent effects in different classes. The empirical laws put forward by Müller and others apply only

to a single class. The measurements analysed are in good qualitative and approximate quantitative agreement with the theory.

The theory also leads to important conclusions regarding dipole induction in radicals forming part of the polar molecule.

A number of special points, such as the attraction of solvent molecules by dipole forces, the influence of anisotropic and polar solvents, the evidence of dielectric constant measurements in water on the structure of Zwitterionen, and the form of the dispersion curve in polar liquids, are discussed in the light of the same theory.

A Study of Sensitized Explosions

I—The Hydrogen Oxygen Reaction Catalysed by Nitrogen Peroxide

By S. G. Foord and R. G. W. Norrish

(*Communicated by T. M. Lowry, F.R.S.—Received April* 27, 1935)

The catalysis of the combustion of hydrogen and oxygen by nitrogen peroxide was first discovered by Dixon,* and later studied quantitatively by Gibson and Hinshelwood,† Thompson and Hinshelwood,‡ and by Norrish and Griffiths.§ Thompson and Hinshelwood found that the ignition temperature of hydrogen was depressed by over 100° C by the addition of less than 0·1% (0·1 mm) of nitrogen peroxide, but that this sensitized ignition was confined within a narrow range of pressures of nitrogen peroxide, the upper limit increasing with rise of temperature and decreasing with increasing pressure, whilst the lower limit was only slightly affected.

Norrish and Griffiths found that, within similar limits of pressure of nitrogen peroxide, rapid reaction but *no explosion* occurred when an axial tube was used for the inlet of gas. This reaction was accelerated by irradiation with light which decomposes nitrogen peroxide to nitric oxide and atomic oxygen, but not by light which is absorbed without producing photochemical decomposition. Without the axial tube, explosions occurred, but the ignition temperature was not affected by irradiation.

---

* Private communication to Mr. Hinshelwood (*see* ref. ‡).

† 'Trans. Faraday Soc.,' vol. 27, p. 559 (1928).

‡ 'Proc. Roy. Soc.,' A, vol. 124, p. 219 (1929).

§ 'Proc. Roy. Soc.,' A, vol. 139, p. 147 (1933).

Thompson and Hinshelwood ascribed the limited range of action of nitrogen peroxide to its dual activity in producing and breaking chains, but in place of their *excited molecules of hydrogen peroxide* Norrish and Griffiths suggested a chain mechanism depending on *atoms of oxygen* generated either thermally or photochemically, and extinguished either by interaction with nitrogen peroxide or by adsorption on the walls of the vessel.

$$\left.\begin{array}{ll}\mathrm{NO_2 = NO + O} & k_0 \\ \mathrm{NO + O_{2_{k_6}}} \underset{k_5}{=} \mathrm{NO_2 + O} & \end{array}\right\}\text{generation of chains}$$

$$\left.\begin{array}{ll}\mathrm{H_2 + O = OH + H} & k_1 \\ \mathrm{H + H_2 + O_2 = H_2O + OH} & \\ \mathrm{OH + H_2 = H_2O + H} & \end{array}\right\}\begin{array}{l}\text{propagation of chains} \\ \quad\text{(Haber mechanism).}\end{array}$$

It was further concluded that the reaction chains were extinguished by reaction of one or other of the free atoms or radicals with the nitrogen peroxide (k_4), or by their adsorption at the wall of the vessel (k_3S). This mechanism leads to the following expression for the velocity of reaction :

$$\frac{d\,[\mathrm{H_2O}]}{dt} = \frac{k k_1 k_0\,[\mathrm{NO_2}]\,[\mathrm{H_2}]}{\{k_3\mathrm{S} + k_4\,[\mathrm{NO_2}] - k_2\}\{k_5\,[\mathrm{NO_2}] + k_1\,[\mathrm{H_2}]\}},$$

k_2 being a factor controlling the branching of chains. If it be assumed that ignition intervenes when the velocity of reaction exceeds a limiting value there are three possibilities, all of which have been realized :

(1) No ignition but only a maximum velocity. The height of the maximum may be increased by an increase of k_0 as is effected by irradiation, but the limits of the catalytic effect remain unchanged.

(2) An upper and lower limit of ignition, as obtained by Thompson and Hinshelwood. This follows when $k_2 < (k_3\mathrm{S} + k_4\mathrm{NO_2})$.

(3) An upper limit, but no lower limit, as obtained by Norrish and Griffiths at higher temperatures. This follows when $k_2 > (k_3\mathrm{S} + k_4\mathrm{NO_2})$.

According to this view ignition is pictured as a thermal process, intervening when the rate of generation of heat exceeds the rate at which it can be conducted away. This view has been criticized recently by Hinshelwood and Williamson* on the grounds that it was not found possible by Norrish and Griffiths to effect a photochemical ignition. They therefore modified the original mechanism of Norrish and Griffiths by introducing a reaction between chain carriers and nitrogen peroxide controlling the branching. It was assumed that chains starting from a particle X would, by reaction with nitrogen peroxide, produce some substance Y capable of giving rise to branching, or alternatively to de-activation by some further reaction with nitrogen peroxide, *i.e.*,

---

* "The Reaction Between Hydrogen and Oxygen." Oxford Univ. Press, p. 83 (1934).

X formed initially.	rate F.
$X + NO_2 = Y + ?$	const. k_1
$Y + H_2 = \alpha X$	const. k_2
$Y + NO_2 \rightarrow$ chain broken.	const. k_3

X destroyed by wall reactions, gas reactions, etc., at a total rate fX.

This scheme gives for the concentration X to which the reaction rate is proportional the expression

$$[X] = \frac{F}{f + k_1 [NO_2] - \dfrac{\alpha k_1 k_2 [H_2] [NO_2]}{k_2 [H_2] + k_3 [NO_2]}}.$$

Limits of ignition are supposed to occur when the positive term ($f + k_1 [NO_2]$) ceases to be greater than the negative term of the denominator, and since the latter grows more rapidly than the former there are two pressures of nitrogen peroxide at which this occurs, in accordance with experiment. As it stands, however, this theory takes no cognizance of the homogeneous reaction which definitely occurs outside the ignition limits,* nor of the rapid reaction which replaces the explosion when the internal surface of the reaction vessel is increased. Neither is any account taken of the fact that at higher temperatures the lower limit of ignition definitely disappears, leaving only the upper. The facts which we have elucidated in the present work show that, contrary to the original view of Gibson and Hinshelwood,* the slow reaction lying outside the limits of ignition is indeed closely connected with the mechanism of ignition, and that it is of exactly the same genus as the explosive reaction within those limits ; both are subject to an induction period. Working at constant temperature it was found that the curve of induction period against pressure of nitrogen peroxide was continuous for the two species of reaction, passing through a minimum and rising to infinite values for $NO_2 = 0$ and $NO_2 = n$, fig. 3. From these results it will be made clear that there is really no distinction in the mode of generation of the slow and explosive reactions ; both intervene sharply at the end of a period of induction, during which the number of chains grows at first slowly and then rapidly to a critical value, as a result of their rate of increase by branching exceeding their rate of extinction. Judged on this basis there is no lower limit of pressure of nitrogen peroxide controlling the reaction limit, and it became of interest to enquire further into the factors which control the ignition point of the mixture. A further and more exhaustive study of the effects of light has confirmed the fact that while the velocity of the slow reaction may be greatly increased by irradiation, and while the induction periods of 200 to 20 seconds of both slow and explosive reaction are reduced to one second or less, there is no detectable effect on the actual limits of ignition. These are, however, materially affected by an increase of internal surface, by the addition of inert gases, or by small quantities of methane, which acts as a powerful anticatalyst.

---

* 'Trans. Faraday. Soc.,' vol. 27, p. 559 (1928).

In the discussion which follows the experimental part of this paper it will be seen that these new results have an important bearing on the mechanism of ignition, and while affording confirmation of the view of Hinshelwood and Williamson that molecules of nitrogen peroxide actually intervene in the branching mechanism, enable a more complete theory of the whole catalytic effect of nitrogen peroxide to be given.

EXPERIMENTAL METHOD

Apparatus—Reactions were carried out in a conical quartz reaction vessel A, fig. 1, fitted at the base end with a plane window, the purpose of which will appear later. This was connected to the remainder of the apparatus by means of a ground joint, and was held horizontally in an electric furnace, the temperature of which was controlled by a thermocouple and automatic regulator, or in certain cases by hand regulation. Any desired gas mixture could be admitted from a mixing vessel B through a wide bore tap C into the evacuated reaction vessel, and observations made of the occurrence of (*a*) explosion, (*b*) slow reaction. Explosions could be detected visually through a small window D in the entry tube, and also audibly by the sharp click which could be heard when ignition occurred.

For the purpose of making up gas mixtures, and for observing phenomena associated with a change of pressure in the reacting system, a special type of Bourdon gauge E has been used. This gauge has been described elsewhere* and is of a particularly robust form, capable of withstanding large pressures (*e.g.*, up to 1 atmosphere) without fracture, while retaining sufficient sensitivity to measure pressure changes to 0·01 mm mercury. The property of immediate importance is that the gauge will withstand the shock of quite vigorous explosions, a fact which has enabled direct observation to be made of the continuous transition from the region of ignition to that of slow reaction, dealt with later in this paper. For accuracy in the measurement of slow reactions the gauge was surrounded by a jacket through which water from a thermostat was circulated.

---

* Foord, 'J. Sci. Instr.,' vol. 11, p. 126 (1934).

The apparatus could be evacuated with a mercury diffusion pump backed by a " Hyvac " pump. For experiments on the photosensitized reaction the radiation from a carbon arc F was directed by means of a quartz lens G into the reaction vessel through the plane end, the infra-red portion of the spectrum being removed by passage through a water filter H. The angle of the cone of light was arranged to coincide with the angle of the reaction vessel in order to obtain a high degree of absorption for a given volume of gas. The ultra-violet portion of the spectrum could be removed at will by the interposition of a quinine filter contained in a parallel-sided glass cell.* Dimensions of the reaction vessel were as follows : length of conical part = 18 cm, tapering from 4·0 cm to 0·4 cm diameter.

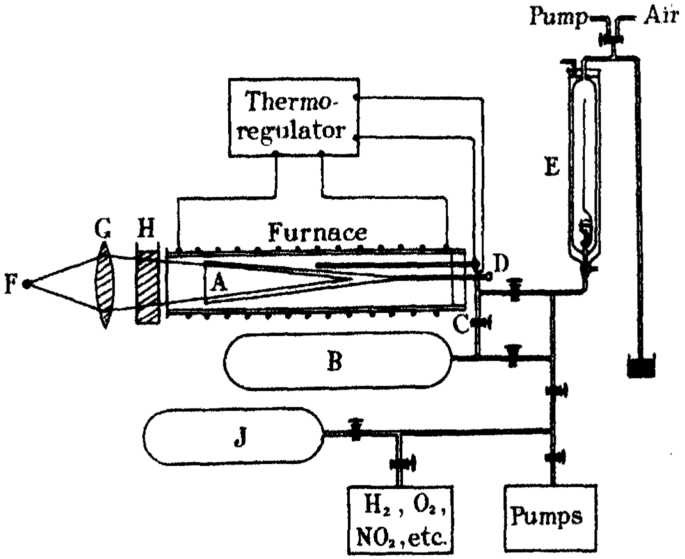


FIG. 1.

---

* Norrish, ' J. Chem. Soc.,' p. 1158 (1929).

Preparation of Gases—Both hydrogen and oxygen were separately prepared by the electrolysis of 10% sodium hydroxide solution saturated with barium hydroxide, and were stored over water. The oxygen was purified from traces of ozone and hydrogen by passage over heated platinized asbestos. Nitrogen and argon used as inert gases were taken from cylinders and used without further purification. Nitrogen peroxide was prepared by heating lead nitrate in a stream of oxygen, condensing at − 80°C, and fractionating in presence of oxygen, the final product being stored in a trap at − 80° C. All gases were dried before use by passage over phosphorus pentoxide.

Practical Methods—For the introduction of accurately known small amounts of nitrogen peroxide, a larger amount was measured off with the calibrated pressure gauge and mixed with oxygen in a separate mixing vessel J, and the requisite amount of this mixture drawn off into the main mixing vessel B, extra oxygen, hydrogen, etc., being then added. The whole was allowed to stand for a minimum time of 10 minutes to obtain a homogeneous mixture by diffusion, and finally admitted to the reaction vessel, which was then connected to the pressure gauge, the latter being filled with a buffer gas to the calculated final pressure. The course of a slow reaction in time could then be followed, or alternatively the length of an induction period, the end of which could be clearly observed, whether terminated by actual ignition or by a slow reaction.

All measurements in the present paper were carried out with a mixture of hydrogen and oxygen in the ratio of 2 : 1, and pressures were adjusted to give a *constant concentration* corresponding to 150 mm pressure at 350° C.

EXPERIMENTAL RESULTS

Effect of Light on Ignition Temperature, etc.—In confirming the negative effect of light on the ignition temperature of a hydrogen-oxygen mixture sensitized by nitrogen peroxide, previously observed by Norrish and Griffiths, a number of measurements have been made of the pressure change in such a system in the neighbourhood of the ignition temperature. Operating at temperatures differing by 1° C, at the higher of which ignition occurred, and at the lower a slow reaction, the same type of reaction occurred in all experiments, whether the reaction vessel was illuminated or not. While the actual ignition temperature was thus unaffected, a very marked reduction of the induction period normally observed was clearly obtained. These results are indicated in the specimen pressure-time curves given in fig. 2, which refer to a nitrogen peroxide pressure of

0·2 mm, for which the ignition temperature was between 356° and 357° C. Curve 1 is a scale representation of the pressure change under conditions which lead to ignition in the dark. During the induction period, which is of the order of 20 seconds in this case, there is a very small decrease in the pressure, about 0·05 mm, after which there is a rapid rise in pressure terminating after ½–1 second in a general ignition, following which the pressure falls as the products of the reaction reassume isothermal conditions. Curve 2 shows the effect of irradiation in the early stages of the induction period, ignition occurring less than one second after exposure to light. Admission of the gas mixture to the reaction vessel with the latter irradiated throughout gives rise similarly to almost immediate ignition.

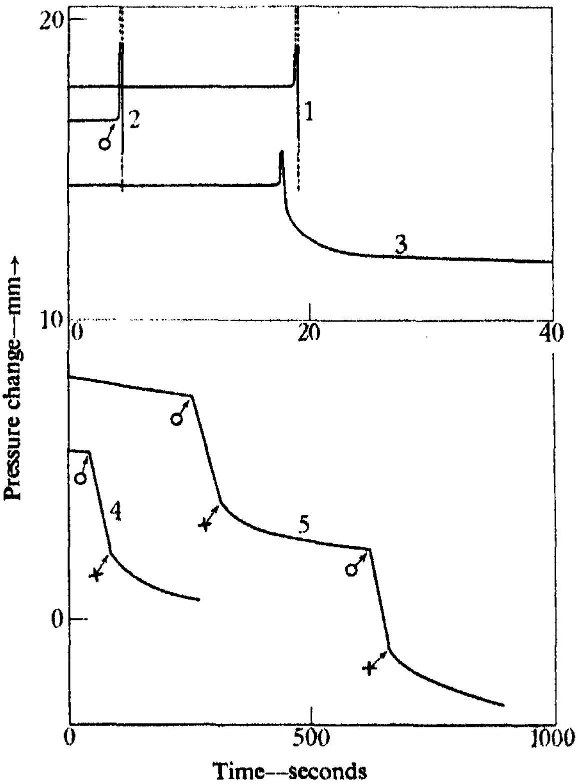


FIG. 2—Pressure time curves at 357° C. Pressure of NO_2 = 0·2 mm. Pressure of $(2 H_2 + O_2)$ = 151 mm. ♂ light on; ↟ light off.

At a temperature just below the ignition temperature, as shown by curve 3, an induction period sharply terminated by a rise in pressure is observed as before. This pressure increase may reasonably be ascribed to the heat resulting from a sudden increase in the reaction rate, but, instead of leading to a general ignition, the reaction slows down to a steady rate, the total fall of pressure due to the initial rapid reaction being of the order of a few millimetres only. The effect of light applied during the induction period is, as before, rapidly to terminate the latter, as shown in curve 4.

A point of particular interest arises in examining the effect of light on the steady reaction finally obtained in curve 3. This is illustrated in curve 5. Exposure of the reacting mixture to light increases the reaction rate tenfold, as has been previously found, but on removing the source of light the residual reaction rate is larger than that obtained before illumination, and only reverts to a value approaching the original after a lapse of a few minutes. An examination of the numerical values of the periods of illumination and of recovery of the original velocity shows that the rate of the slow dark reaction is intimately bound up with the degree of dissociation of the NO_2 into NO.

The Induction Period—The clearly defined nature of the induction period, even where actual ignition does not occur, renders it of considerable value in determining the nature of the processes leading to explosion. A number of investigations has been made of the variation of the induction period with the concentration of nitrogen peroxide and certain other variables, and, as will be seen, the results of these measurements provide considerable evidence that the mechanism of ignition is in the end a thermal process.

Fig. 3 shows the variation of induction period with the concentration of nitrogen peroxide at a constant temperature of 357° C. It is seen that the induction period passes from a very large value at zero concentration of nitrogen peroxide through a minimum value, and rapidly becomes very large again as the amount of nitrogen peroxide is increased. Between the two limits of concentration A and B, complete ignition of the mixture resulted, but on either side of these limits, whilst no directly visible explosion occurred, pressure-time curves of the type shown in curve 3 of fig. 2 were obtained ; in these regions the maximum reaction rate following immediately after the induction period became smaller the farther one receded from the limiting concentrations A and B.

A parallel series of measurements with the reaction vessel irradiated showed in a striking manner the shortening of the induction period by light, and the experimental points are indicated by the black circles at the foot of fig. 3, all induction periods of one second or less being marked as one second. The limiting concentrations of nitrogen peroxide between which ignition occurred were found to be exactly the same as in the dark reaction.

With a mixture containing 0·4 mm of nitrogen peroxide (*i.e.*, within the ignition region), experiments with smaller light intensities gave a progressive reduction of the induction period from its normal dark value as the intensity was increased by means of an adjustable stop, giving the results shown in fig. 4. The significance of this is discussed later in connexion with the mathematical theory of the induction period.

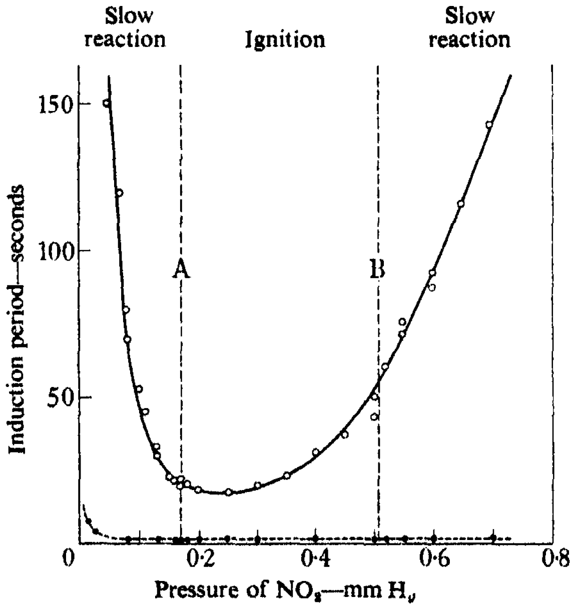


FIG. 3—Induction periods at 357° C. $(2H_2 + O_2) = 151$ mm. ——○ dark; - - ● - - light.

A similar series of measurements to that given in fig. 3 for the dark reaction was made at a temperature of 390° C, and is illustrated in fig. 5. The increase of temperature has brought about a reduction of the length of induction periods, but the shape remains of the same form. The range of concentrations of nitrogen peroxide, inside which complete ignition occurs, is considerably widened, as is to be expected from the curves obtained by Thompson and Hinshelwood; but the position of minimum induction period remains the same as at the lower temperature.

Other measurements were made showing the effect of the addition of certain foreign gases on the induction period, using a fixed temperature of 357° C and a fixed pressure of 0·2 mm of nitrogen peroxide. The results for argon and nitrogen and for methane are shown in Table I. The first

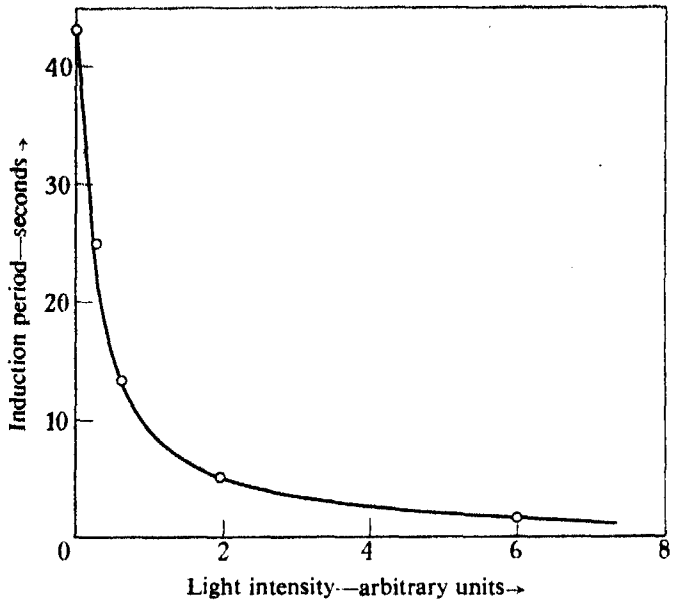


FIG. 4—Induction periods at 157° C under irradiation. Pressure of NO_2 = 0·4 mm; pressure of $(2H_2 + O_2)$ = 151 mm. The full line represents the theoretical curve.

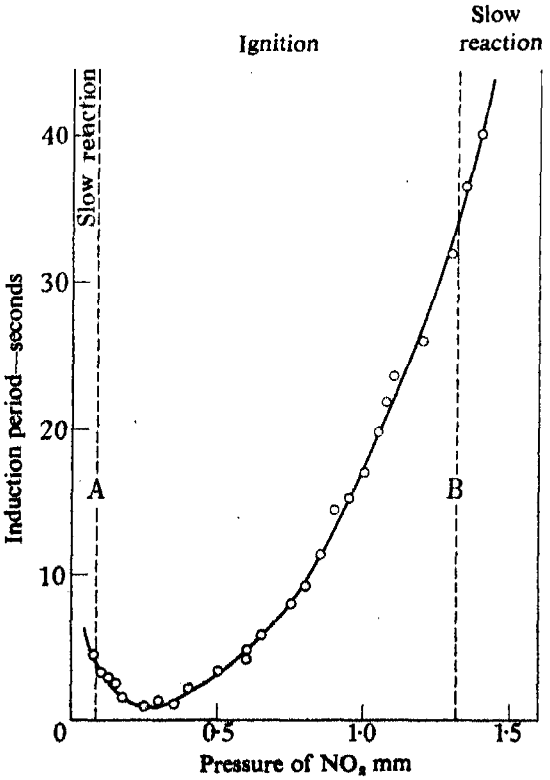


FIG. 5—Induction periods at 390° C. $(2H_2 + O_2)$ = 160 mm.

TABLE I—EFFECT OF FOREIGN GASES ON INDUCTION PERIOD AT 357° C
Pressure of $(2H_2 + O_2)$ = 151 mm. Pressure of NO_2 = 0·2 mm.

Argon		Nitrogen		Methane	
Argon pressure mm	Induction period secs	N_2 pressure mm	Induction period secs	CH_4 pressure mm	Induction period secs
0	25*	0	27*	0·00	20*
10	24½*	50	27*	0·11	26*
20	25*	100	32	0·19	31½
30	24½*	150	44	0·22	29½
50	24½*			0·41	39
120	24½*			0·57	44½
200	30			0·76	60
				0·95	79
				1·14	98
				1·34	>500

* Represents an ignition; unmarked figures represent slow reaction.

two gases have only a small effect, but in sufficient quantity, >100 mm, tend to increase the induction period and ultimately to convert the ignition into a slow reaction. The effect of methane, however, is of an entirely different order to that of the argon and nitrogen. At pressures greater than 1·5 mm it completely inhibits the reaction ; a pressure of methane less than 0·2 mm is sufficient to convert the ignition into a slow reaction, while a pressure of 1·0 mm increases the induction period more than three-fold. Thus methane is found to be a powerful anticatalyst for the sensitized ignition of hydrogen, and we have the interesting phenomenon of the suppression of combustion by a trace of a second combustible substance.

In fig. 6 is shown the effect of varying the pressure of oxygen. Here again two ignition limits are disclosed, but the curve of induction periods runs continuously from the region of slow reaction to the region of ignition.

DISCUSSION

The experiments described in the foregoing sections establish the following facts :

(1) Nitrogen peroxide acts as a catalyst for the homogeneous combination of hydrogen and oxygen, and is effective both for the slow and explosive reactions. Its effect passes through a maximum and falls rapidly to zero as its concentration is increased.

(2) In the middle of this region of catalysis, ignition may occur between certain sharply defined limits. Both slow and explosive reactions are of the same genus, each being subject to an induction period. The plot of the induction periods against increasing pressure of nitrogen peroxide is continuous for both slow and explosive reactions, and passes through a minimum corresponding to the maximum of catalysis.

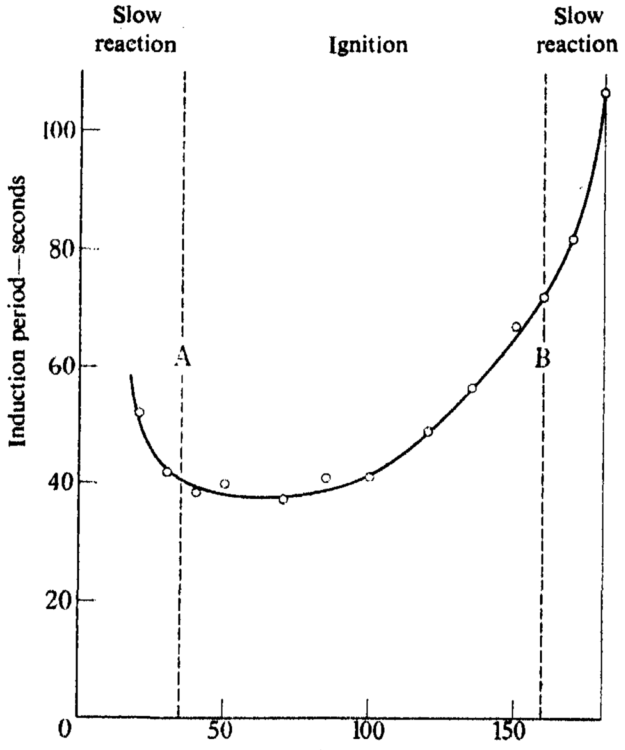


FIG. 6—Effect of oxygen at 357° C. Pressure of NO_2 = 0·25 mm; pressure of hydrogen = 151 mm.

(3) Immediately outside the ignition limits small but sharp pressure increases mark the end of the induction period and precede the onset of slow reaction. The magnitude of these pressure effects decreases as we pass away from the limits of ignition.

(4) The velocity of the slow reaction catalysed by nitrogen peroxide is markedly increased by irradiation by light which effects the photo-decomposition of the nitrogen peroxide. This increase in velocity may persist for some time after the light is extinguished, but the velocity ultimately returns to the normal " dark " value.

(5) The induction periods of both slow and explosive reactions are greatly reduced by irradiation, being decreased from values between 20 and several hundred seconds to less than one second over nearly the whole range of catalytic activity of the nitrogen peroxide.

(6) The limits of the ignition region occurring within the range of catalytic activity of nitrogen peroxide could not be influenced by irradiation.

(7) These limits are nevertheless markedly affected by a change in the internal surface of the reaction system, and by a sufficient increase in the surface to volume ratio the ignition reaction may be suppressed altogether.

(8) The limits are also affected in the direction of partial suppression of reaction by an increase of the pressure of oxygen, or by the addition of inert gases such as nitrogen or argon.

(9) The effect of small traces of methane is exceedingly great, quantities as small as 1 part in 10^3 being sufficient to suppress the ignition, while somewhat larger quantities suppress the slow reaction entirely.

From the facts enumerated above, it may be concluded that the combination of hydrogen and oxygen occurs by a chain mechanism, starting initially from a small number of centres, generated either thermally or photochemically from molecules of nitrogen peroxide. These chains, besides being self-propagating, may also multiply in number by a branching mechanism; they are also subject to extinction by reaction with nitrogen peroxide in the body of the gas, and also by absorption of their active carriers at the surface of the reaction vessel. The two factors constitute the predominant de-activating influences on the reaction in the presence of nitrogen peroxide, but, by the addition of traces of methane, a new and powerful chain-breaking mechanism is introduced.

Immediately after the introduction of the reactive mixture of hydrogen, oxygen, and nitrogen peroxide into the dark reaction system, the velocity of reaction is undetectably small, because the number of centres is extremely low. There follows a period of induction, during which the number of centres is exponentially increasing by the branching mechanism until the velocity of reaction becomes measurably large. The rate of growth of the reaction during the induction period is dependent upon the "effective branching factor"—that is, the difference between the probability of branching and the probability of extinction of chains, and if the effective branching factor be small, long induction periods measured in seconds or minutes may result, during which there is practically no detectable reaction. At the end of this induction period the number of centres increases very rapidly and the velocity becomes measurable. It is important to realize, as will be shown mathematically below, that all the characteristics of the induction period and its variation with the concentration of reactants can be accounted for by the above theory, which postulates that there is a positive rate of growth of reaction centres in the body of the gas.

Conditions of Ignition—The fact that the induction periods follow a uniform curve over the whole region of catalysis by nitrogen peroxide, fig. 3, shows that the development of both slow and ignition reactions occurs by identical mechanisms. It is clear that for both types of reaction the effective branching factor during the induction period must be positive. In spite of this, however, the number of reaction centres will not increase exponentially to infinity, for when their number (n) becomes very large, a new process of de-activation will intervene owing to their self-neutralization. Since this process must depend upon n^2 it will be negligible during the induction period when n is small, and will only assume the role of a controlling factor at the end, when n is large. The result is that n rises to a large finite value instead of infinity.

If θ be the rate of production of chains by the primary process, and ϕ be the *effective* branching factor (*i.e.*, the difference between the probabilities of branching and de-activation exclusive of self-neutralization), while δ is the factor controlling self-neutralization, then

$$\frac{dn}{dt} = \theta + \phi n - \delta n^2. \qquad (1)$$

The length of the induction period is then obtained from the solution of equation (1). Neglecting the small term δn^2 we obtain

$$t = \frac{1}{\phi} \log \left\{1 + \frac{n\phi}{\theta}\right\}. \qquad (2)$$

The plot of n against t is shown in the right-hand full curve of fig. 7 for suitable arbitrary values of θ and ϕ; it is apparent that so long as $n\phi/\theta$ remains small compared with unity there is a linear increase of n with t, while for values of $n\phi/\theta$ considerably greater than unity, as at the end of the induction period, n increases exponentially with t. In view of this final rapid increase in n it is sufficient to take some large arbitrary value n as representing a critical concentration to which we measure all our induction periods; any further growth of n beyond this value will occur in so short a time that it will make no appreciable difference to the induction period. Thus the equation for the induction period τ becomes

$$\tau = \frac{1}{\phi} \log \left\{1 + \frac{n_c\phi}{\theta}\right\}. \qquad (3)$$

At the end of the induction period the factor δn^2 comes into prominence owing to the rapid increase in the number of chains, and the system will finally tend to a potential equilibrium, given by equating dn/dt to zero in equation (1). This gives for the equilibrium value of n,

$$n_e = \frac{\sqrt{4\delta\theta + \phi^2} + \phi}{2\delta}, \tag{4}$$

as shown qualitatively by the dotted portion of the curve in fig. 7. The magnitude of this value of n_e will vary with ϕ, the effective branching factor, which is a function of the concentration of nitrogen peroxide, and if ϕ^2 assumes large values in comparison with $\delta\theta$ will be mainly fixed by the expression

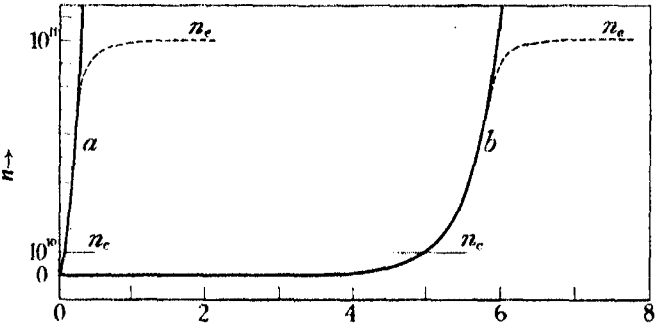


FIG. 7—a light; b dark.

$$n_e = \frac{\phi}{\delta}. \tag{5}$$

Now it is important to realize that the surface de-activation factor which is contained in ϕ will vary somewhat from point to point in the reaction system, as a function of the distance to the surface. Thus ϕ will vary similarly from point to point, and there will therefore be some element of volume for which ϕ is a maximum. In this element of volume n_e will reach its greatest value, and this region will be the seat of the greatest intensity of reaction.

If the intensity of reaction be great enough, the rate of production of heat within the element will be greater than the rate at which it can be conducted away, the process will become adiabatic, and ignition will spread out from that point. The occurrence of ignition is thus seen, according to equation (5), to be dependent on the attainment of a limiting value of ϕ at some point in the reaction system, and since ϕ is a function of the concentration of nitrogen peroxide c, this may occur for more than one value of the latter. Experimentally there are found to be two ignition limits, and therefore any theory of the mechanism must yield ϕ as a quadratic function of c.

In view of the thermal mechanism of ignition given above, it is instructive to examine the character of the curves for the slow reaction shown in fig. 2. In the region just outside the ignition limit it is seen that the induction period is terminated by a sharp pressure increase followed by a fast reaction, which slows up considerably to a steady value in the course of a minute or so (*see* curve 3 of fig. 2). Since the reaction itself occurs with a fall of pressure, the sharp increase of pressure terminating the induction period is clearly a heat effect, and marks the tendency of the reaction to become adiabatic at this point. The particular form of the curve leads one to believe that we are here witnessing an incipient thermal ignition in some favourable volume element, which rapidly dies away as the reactants in the element are consumed. The process once more becomes isothermal and a steady reaction velocity conditioned by equation (4) ensues.

Objections have been raised to any thermal theory of ignition on the grounds that transition from slow reaction to ignition is very sharp, and that at higher pressures or with increased surface faster reaction rates may be observed than those which normally lie immediately adjacent to the ignition region at ordinary pressures. These objections disappear when it is remembered that owing to the wide variation of ϕ from point to point in the reaction system, it is only necessary for the critical value of n_e to be exceeded in some small volume element in order to obtain ignition. The value of n_e in other elements may fall far below this, and a " volume average " will thus give a very small " over-all " reaction velocity. If, however, ϕ be decreased by increase of surface, or δ be increased by increase of pressure (ternary collision effect), it is readily seen from equation (4) that an increase in θ can greatly raise the average value of n_e over the whole vessel, even though the limiting value of n_e is not approached at any point in the system. This is what occurs when an axial tube is introduced into the reaction vessel, for ϕ is decreased due to the increase of surface dimensions, and θ steadily increases with increase in pressure of nitrogen peroxide, as we pass into the region which previously gave ignition. The system is, however, now stable, since ϕ is small, and at no point does n_e reach the peak value required for ignition. Nevertheless, the average value of n_e taken over the whole vessel may be considerably greater than heretofore, and relatively high overall reaction velocities may become possible.

The Effect of Irradiation—From the fact that the ignition limits are not measurably affected by irradiation, it is clear that the light can have no *immediate* effect on the effective branching factor ϕ. Its effect in shortening the induction period is to be ascribed to a marked change in the value of θ. If we assume that for the dark reaction $\phi/\theta = 10^{-5}$ and that θ is increased by a factor of 10^6 by irradiation, then, taking n_0 (the initial number of centres) as 10^4 and $n_c = 10^{10}$ per cc, we may calculate from equation (3) that the induction period is reduced by 100-fold in the light. This is of the same order of magnitude as is observed experimentally. If now we take $\delta/\phi = 10^{-10}$ as a probable value, we find from equation (4) that the equilibrium concentration of centres n_e is *ca.* 10^{11}/cc, and is only increased by about 1% by irradiation. Thus, in spite of the great shortening of the induction period, the ignition temperature will not be appreciably affected unless much greater intensities of light are used. The curves of fig. 7 have been drawn assuming the above numerical values.

Turning to the slow reaction, however, we find the apparent anomaly that the velocity may be increased as much as tenfold by irradiation, and this should indicate a corresponding increase in n_e, and therefore in ϕ by equation (4). The explanation of this probably depends on the observation that immediately after the end of the induction period there is a rapid reduction of the velocity of the slow reaction to a steady value some 100-fold to 10-fold smaller than its initial rate. (*See* curve 3, fig. 2, which is typical of many others we have obtained.) It appears probable, therefore, that the products formed in the first " burst of reaction " may reduce the effective branching factor ϕ by a factor of 10 to 100, and this must effect a stabilization of the mixture against ignition. Assuming this to be true, and assuming the numerical values of ϕ/θ and δ/ϕ quoted above, it will be seen that $\delta\theta$ in equation (4) is now not negligible compared with ϕ^2. Assuming, as above, that ϕ is reduced by the factor 10^{-2} following the onset of the slow reaction, and that as before θ is increased by a factor 10^6 by irradiation, we may calculate that the magnitude of n_e is now reduced to 10^9 centres/cc in the dark, and increases to 10^{10}/cc by irradiation. Thus the velocity of slow reaction should be increased by about 10-fold, as was found experimentally.

The values of ϕ/θ and δ/θ, n_c and n_0, etc., assumed above conform satisfactorily to all the experimental results. They are quoted to show that the present theory can give an adequate description of the facts, and, while in themselves not improbable values, they are not suggested as absolute magnitudes.

The variation of induction period with the intensity of irradiation may be investigated by equation (3). If the value of θ for the dark reaction be represented by a, and its value in the volume element at the point of focus during irradiation by $(a + b\mathrm{I})$, where I is the light intensity, then for a constant value of ϕ, equation (3) takes the form

$$\tau = k \log \left\{1 + \frac{k'}{a + n\mathrm{I}}\right\}. \tag{3a}$$

If we assign appropriate values to the constant, τ may be plotted as a function of I, and we obtain the full curve shown in fig. 4. It is seen that the experimental points also marked in the figure are in satisfactory agreement with the form of the theoretical curve.

Branching of Chains—As has been mentioned, it was originally assumed by Norrish and Griffiths that the action of nitrogen peroxide is confined to the processes of starting and stopping of chains, and that it is without effect on the branching mechanism. Hinshelwood and Williamson, however, expressed the opinion that in addition the branching must be directly dependent on the pressure of nitrogen peroxide. This was necessary in order to explain the supposed lower limit of catalysis by nitrogen peroxide without invoking the mechanism of a thermal ignition, which appeared to be excluded by their interpretation of the results of Norrish and Griffiths. The discovery that in reality there is no appreciable lower limit to the region of catalysis by nitrogen peroxide, and the conclusion from the present data that the ignition is in effect a thermal process, removes the original grounds for the assumption of Hinshelwood and Williamson; we have nevertheless found it desirable to retain their assumption, for it appears to be essential to the explanation of the exact mathematical form of the experimental curve of induction periods shown in fig. 3.

The Effect of Added Substances—From Table I it is seen that the addition of argon in quantities up to 200 mm has but a small effect on the induction period, but that in sufficient quantity it converts the ignition into a slow reaction. The same is true for nitrogen, but its effect on the induction period is somewhat greater. From their effect in quenching the ignition we may conclude that the action of these gases is in the nature of a " three body effect." The self-neutralization of reaction centres, which, as we have seen, must become appreciable at the end of the induction period, will be subject to a stabilization by a ternary collision, and thus the factor δ will be proportional to the pressure, and increased by the addition of a foreign gas. By equation (4) an increase in δ will depress the value of n_e, the equilibrium value to which the concentration of reaction centres tends, and if this depression be great enough a quenching of the thermal ignition will obviously result. The effect of nitrogen in increasing the induction period indicates also that it has a small influence in reducing the effective branching factor ϕ.

The effect of methane is of an entirely different order. The action of very small quantities in quenching the ignition and greatly increasing the induction period is only to be explained by a marked decrease in the value of the factor ϕ, as the result of a powerful de-activating effect upon the reaction chains. Such an effect will arise if the methane molecule reacts preferentially with some component of the hydrogen oxygen chain, and forms a product which takes no further part in chain propagation, and is possible since the temperature of the system, 357° C, is well below that at which the combustion of methane itself is propagated by a chain mechanism.

The effect of varying the oxygen concentration shown in fig. 6 provides further evidence for the continuity of the slow and ignition reactions. With too much or too little oxygen the mixture ceases to ignite, as is well known, but in the region outside these ignition limits, the slow reactions are clearly subject to the same process of development as in ignition. We cannot escape the conclusion that here, as in the experiments of fig. 3, there is a positive effective branching factor over the whole region of reaction, and that the incidence of ignition is a thermal effect, as has already been explained above.

THEORY OF REACTION KINETICS

The following assumptions are made in developing expressions for the rate of reaction, explosion limits, induction period, etc.:

(*a*) Chains are started by oxygen atoms, *e.g.*, by reaction with hydrogen to give H + OH.

(*b*) Chains are terminated by a suitable collision between a nitrogen peroxide molecule and the chain, or by a surface reaction.

(*c*) A different type of collision between a nitrogen peroxide molecule and a chain can result in the production of an oxygen atom, without necessarily stopping the original chain, resulting in a net branching if the oxygen atom starts another chain or chains.

(*d*) Oxygen atoms are produced by primary processes dependent upon the nitrogen peroxide, which for simplicity we shall first suppose to be represented by

$$NO_2 = NO + O \qquad (k_1). \qquad \text{(I)}$$

They are also produced by the branching mechanism described under (*c*) above. They may be removed in the gas phase by reaction with nitrogen peroxide or hydrogen,

$$NO_2 + O = NO + O_2 \quad (k_2) \tag{II}$$

$$O + H_2 = OH + H \quad (k_3), \tag{III}$$

and at the surface by adsorption, *e.g.*,

$$O + \text{surface} = \tfrac{1}{2}O_2 \quad (k_4). \tag{IV}$$

There results at every moment a dynamic equilibrium in which the concentration of oxygen atoms is fixed by the above processes.

Oxygen Atom Equilibrium—If

c = concentration of nitrogen peroxide,

n = concentration of reaction centres (number of chains/cc),

then the rate of production of oxygen atoms by the branching mechanism may be represented by αnc.

If, further, h is the concentration of hydrogen, it follows that the instantaneous concentration of oxygen atoms is given by

$$[O] = \frac{k_1 c + \alpha nc}{k_2 c + k_3 h + k_4}. \tag{6}$$

Chain Development—If the rate of termination of chains by nitrogen peroxide be represented by βnc, and γn represents the rate of termination by the surface, excluding the removal of oxygen atoms which is already taken account of by k_4, then the rate of production of chains is given by

$$\frac{dn}{dt} = k_3 h\,[O] - \beta nc - \gamma n. \tag{7}$$

Substituting from equation (6) we obtain

$$\frac{dn}{dt} = \left(\frac{k_1 c + \alpha nc}{k_2 c + k_3 h + k_4}\right) k_3 h - \beta nc - \gamma n \tag{8}$$

$$= \frac{k_1 k_3 hc}{k_2 c + k_3 h + k_4} + \left\{\frac{\alpha k_3 hc}{k_2 c + k_3 h + k_4} - \beta c - \gamma\right\} n. \tag{9}$$

Writing

$$\theta = \frac{k_1 k_3 hc}{k_2 c + k_3 h + k_4}, \tag{10}$$

and

$$\phi = \frac{\alpha k_3 hc}{k_2 c + k_3 h + k_4} - \beta c - \gamma, \tag{11}$$

equation (9) may be written

$$\frac{dn}{dt} = \theta + \phi n. \tag{12}$$

This is identical with equation (1) if we neglect the term δn^2, which only becomes of importance at the end of the induction period. The expressions (10) and (11) give the forms of θ and ϕ in terms of the nitrogen peroxide concentration c; these, after simplification, may be substituted in equation (3) to obtain the variation of the induction period τ with c. If for the moment we consider the rate of surface termination of chains to be small compared with their termination by nitrogen peroxide, we obtain

$$\phi = \frac{c}{k_2c + k_3h + k_4}\{\alpha k_3h - \beta(k_2c + k_3h + k_4)\} \tag{13}$$

$$= \beta\frac{c}{c + k_5}(k_6 - c), \tag{14}$$

where

$$\left.\begin{aligned} k_5 &= \frac{k_3h + k_4}{k_2}, \\ k_6 &= \frac{\alpha k_3h - \beta k_3h - \beta k_4}{k_2}, \end{aligned}\right\} \tag{15}$$

and

$$\theta = \frac{k_1k_3hc}{k_2(c + k_5)}. \tag{16}$$

Substituting in equation (3),

$$\tau = \frac{c + k_5}{\beta c(k_6 - c)}\log\left\{1 + \frac{\beta n_0(k_6 - c)}{k_1k_3h}\right\}. \tag{17}$$

This expression gives with the introduction of suitable values for constants a curve of the form shown by the full line in fig. 8, which is closely similar to the experimental curve shown in fig. 3. It is seen that the curve is asymptotic to the ordinate $c = 0$, and that no lower limit of nitrogen peroxide is indicated below which the catalytic effect ceases to exist. The inclusion of the term γ, representing the extinction of reaction chains at the surface, which we previously neglected, introduces such a limit and has the effect of contracting the whole curve without appreciably altering its form, as is shown by the dotted curve in fig. 8 for a particular value of γ. We may therefore conclude, in view of the fact that the experimental curve of fig. 3 is also nearly asymptotic to the ordinate, that γ is in effect very small, and that practically the whole effect of the surface is exerted through the factor k_4 controlling the equilibrium of oxygen atoms.

Mechanism of Starting of Reaction Chains—The actual primary mechanism by which the oxygen atoms required for the starting of reaction chains are produced is somewhat immaterial to the development of the thermal reaction kinetics, and while the reaction (I) assumed above is probably right for the photochemical process, the reaction

$$NO + O_2 = NO_2 + O, \qquad (V)$$

cannot be excluded as a primary source of oxygen atoms in the thermal reaction. The substitution of reaction (V) for reaction (I) involves the replacement of the term k_1c by k_1c^2t in equation (6) and its derivatives, since the nitric oxide is produced by the slow thermal decomposition of nitrogen peroxide at a rate which does not appreciably change the value of c during the normally observed induction period. Solving, as before, for the development of reaction centres with time, we obtain for the induction period the approximate relation

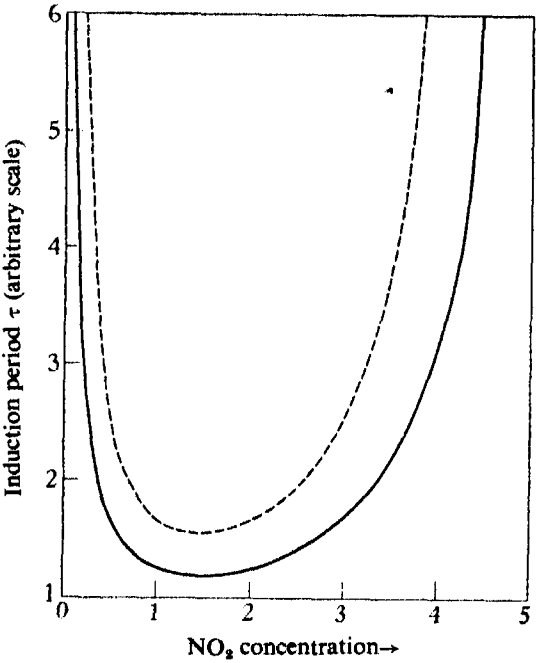


FIG. 8.

$$\tau = \frac{1}{\phi} \log \left\{1 + \frac{n_c\phi^2}{\theta}\right\}. \qquad (18)$$

Since, as we have seen in connexion with equation (3), a sharp induction *period is conditioned by the term inside the logarithm being considerably* greater than unity, the shape of this curve will be determined mainly by the expression $1/\phi$, and will therefore be similar to that of equation (17) after substitution for ϕ in terms of c.

The Nature of the Chain Mechanism—The propagation of reaction chains through mixtures of hydrogen and oxygen has been a matter of study from several points of view. Comparing the results of Marshall,* Bates, and Salley,† and Frankenburger and Klinkhardt‡ at low temperatures with those of Haber§ and his co-workers at high temperatures, we conclude that the following mechanism seems to possess the greatest probability :

$$H + O_2 = HO_2 \tag{VI}$$

$$HO_2 + H_2 = H_2O + OH \tag{VII}$$

$$OH + H_2 = H_2O + H. \tag{VIII}$$

At temperatures approaching the ignition point the self-accelerative character of the reaction manifested in the induction periods indicates the existence of a branching mechanism, and we may assume that occasionally reaction (VI) may take the alternative form

$$H + O_2 = HO + O, \tag{IX}$$

followed by

$$O + H_2 = HO + H, \tag{X}$$

involving a multiplication of reaction centres.

Now at temperatures far below the ignition point of hydrogen and oxygen, at which the nitrogen peroxide is an effective catalyst, it is clear that the branching mechanism cannot be operative, otherwise the mixture would ignite in the absence of the catalyst. We must therefore conclude that in the presence of nitrogen peroxide a new branching mechanism comes into prominence, and this provides further justification for the suggestion of Hinshelwood and Williamson which we have adopted above. The chain process itself is highly exothermic, and, without specifying any particular reaction, it is easy to see that the probability of the reaction

$$NO_2 = NO + O$$

---

* 'J. Amer. Chem. Soc.,' vol. 54, p. 4460 (1932).

† 'J. Amer. Chem. Soc.,' vol. 55, p. 110 (1933).

‡ 'Trans. Faraday Soc.,' vol. 27, p. 431 (1931).

§ Bonhoeffer and Haber, 'Z. phys. Chem.,' A, vol. 137, p. 263 (1928); 'Z. angew. Chem.,' vol. 42, p. 475 (1929). Farkas, Haber, and Harteck, 'Naturwiss.,' vol. 18, p. 266 (1930); 'Z. Electrochem.,' vol. 36, p. 711 (1930).

is greatly increased by collision of the molecule of nitrogen peroxide with some highly energized chain carrier, or alternatively by its entrance as a third body into one of the exothermic links of the chain.

This generation of oxygen atoms constitutes a branching factor which is obviously proportional to the product of the nitrogen peroxide concentration and the number of chains, as has been assumed above. We have seen that kinetically nitrogen peroxide also exerts an anticatalytic effect in extinguishing the reaction chains. This may take place by the highly efficient reaction

$$NO_2 + O = NO + O_2,$$

and also by the reaction of other chain carriers with nitrogen peroxide, as, for example,

$$NO_2 + OH = HN\dot{O}_3. \qquad \text{(XI)}$$

CONCLUSION

The results described above indicate that the generally accepted view that ignition occurs when the rate of branching of chains just exceeds their rate of extinction must be modified, since our observations show that this condition is also fulfilled during the induction period of both slow and ignition reactions. At the end of the induction period, the number of reaction centres tends rapidly to an equilibrium value due to the processes of self-neutralization coming into prominence, and ignition finally occurs only if this concentration of centres in some favourable volume element exceeds a limiting value. Under these circumstances the process locally becomes adiabatic, and inflammation may spread throughout the whole system.

A complete theory of the photochemical and thermal processes involved in the combustion of hydrogen and oxygen in the presence of nitrogen peroxide can be built up on the assumption that the effective catalyst is the oxygen atom, and it may be concluded with some certainty that the nitrogen peroxide plays a triple role in influencing the mechanism of origination, branching, and extinction of chains.

The authors desire to express their thanks to the Royal Society and to Imperial Chemical Industries for grants which have made this work possible.

SUMMARY

In extending earlier work on the catalysis of the combustion of hydrogen and oxygen by traces of nitrogen peroxide, it has been found that the

region of ignition which occurs between narrow limits of concentration of the catalyst is bounded both on the upper and lower limits by regions of easily measurable slow reaction.

The velocity of this slow reaction is considerably increased by light which generates oxygen atoms from the nitrogen peroxide, but the ignition limits are unaffected.

Both slow and ignition reactions are subject to induction periods, which in the slow reaction are terminated by a sharp pressure increase followed by a reaction whose velocity rapidly diminishes to a steady value. At constant temperature, the curve of these induction periods runs continuously from the region of slow reaction to the region of ignition, and passes through a minimum corresponding to the maximum of catalytic effect. Judged on this basis there is a definite upper limit to the catalytic effect of nitrogen peroxide for which the induction period is infinite, but the lower limit, if it exists at all, occurs at a pressure of nitrogen peroxide which is less than 0·01 mm at 350° C.

The induction periods are reduced by irradiation from values of 200–20 seconds to values less than one second over nearly the whole range of catalytic effect of the nitrogen peroxide.

The effect of foreign gases on the induction period has been investigated.

It is concluded that :

(*a*) The branching initially exceeds the extinction of chains over the whole region of catalysis.

(*b*) An equilibrium concentration of reaction chains is ultimately imposed on the slow reaction by the self-neutralization of chains, and by an increase in their extinction due to products of reaction.

(*c*) Ignition occurs as a thermal effect due to the process becoming approximately adiabatic when the number of chains exceeds a critical value.

(*d*) Nitrogen peroxide exercises a triple role in affecting the origin, branching, and extinction of chains.

A mechanism for the process based on the production and removal of oxygen atoms by the catalyst is suggested.

A generalization relative to the process of ignition in general is indicated.

The Continuous Absorption Spectrum of Hydrogen Bromide

By C. F. GOODEVE and A. W. C. TAYLOR, The Sir William Ramsay Laboratories of Inorganic and Physical Chemistry, University College, London

(*Communicated by F. G. Donnan, F.R.S.—Received April* 30, 1935)

Some molecules absorb light continuously at the high-frequency (high energy) end of a band system, while others show a similar continuous absorption without any related band system. Molecules of the halogens belong to the first type; their upper electronic levels are stable and show a series of vibrational sub-levels. From an analysis of the fine structure of the band spectrum, it is possible to calculate the course of the potential energy-internuclear distance curve for the upper state. Furthermore, the convergence limit of the band system corresponds exactly to the heat of dissociation into one excited and one normal atom. A knowledge of the excitation energy of the products thus leads to the normal heat of dissociation of the molecule.

It has sometimes been concluded from the absence of band systems with molecules of the second type, such as the hydrogen halides, that no minima exist in the potential energy curves of the upper state in these cases. Furthermore, the coincidence between the position of the convergence limit of a band system and the " threshold " of continuous absorption had led to the conclusion that this threshold with molecules of the second type corresponds also to a heat of dissociation. This simple conclusion has not been substantiated by recent theoretical developments; but these, on the other hand, give us a method of calculating the course of the upper potential energy curve from the extinction coefficients of absorption. In the present investigation the values of this coefficient for hydrogen bromide have been measured accurately over a wide range of conditions, and the approximate upper potential energy curve has been calculated.

The absorption spectrum of HBr has been studied qualitatively by Coehn and Stuckardt,[*] Warburg,[†] and Tingey and Gerke.[‡] Dutta[§]

---

[*] 'Z. phys. Chem.,' vol. 91, p. 722 (1916).

[†] 'S.B. preuss. Akad. Wiss.,' vol. 76, p. 314 (1916).

[‡] 'J. Amer. Chem. Soc.,' vol. 48, p. 1838 (1926).

[§] 'Z. Physik,' vol. 77, p. 404 (1932).

measured some values of the extinction coefficient, but his results are not in agreement with those of Tingey and Gerke nor with our own. Moreover, his conclusions based on an extrapolation of the extinction coefficient to zero are unjustified, since this coefficient decreases exponentially with decreasing frequency.

OPTICAL MEASUREMENTS

The extinction coefficient is here defined in the usual manner by the expression

$$\alpha = \log_{10}\frac{I_0}{I_T} \times \frac{1}{p.l.}, \tag{1}$$

TABLE I

	Light source	Absorption cell	Range of extinction coefficient	Method of photometry	Spectrograph
I	Hydrogen discharge tube	1 cm fused quartz	10^{-1} to 2×10^{-4}	Single cell, rotating sector	Hilger small quartz E.370
II	Iron-tungsten spark	8 cm fused quartz	3×10^{-3} to 7×10^{-5}	Twin cells, Hilger Spekker divided beam photometer	Hilger medium quartz E.3
III	Fe spark	138 cm fused quartz	2×10^{-4} to 2×10^{-6}	Single cell, rotating sector	Bellingham and Stanley medium quartz
IV	Hydrogen discharge tube	3355 cm glass with quartz end-pieces	4×10^{-6} to 5×10^{-8}	Single tube, rotating sector	Zeiss constant deviation quartz, also Hilger quartz Raman E.420

where I_0 and I_T are the intensities of the incident and transmitted beams, p, the pressure of the gas in millimetres Hg, and l the length of the absorbing column in centimetres.

The measurements have been made by a photographic method. The intensity of the light transmitted by the gas was compared with that of light of the same initial intensity, reduced by known amounts by means of a rotating sector, or other standard device. The " match point," the frequency at which the photographed spectra were of equal blackening, was usually determined visually. Pressures were measured directly on a mercury manometer.

Four different optical systems were used to cover the wide range of optical thickness (= pressure × length) necessary to avoid possible complications due to high pressures. These systems are summarized in Table I.

In the first and third ranges, as given in Table I, the comparison spectra were taken immediately before and immediately after the absorption spectrum and through the same tube, the gas being removed by condensation in a liquid air trap. The constancy of the light source could thus be confirmed. Schumann plates were used at frequencies greater than 45,000 wave numbers. In the second range, as given in Table I, a " Spekker " photometer* was used.

With the 3355 cm tube a slightly different technique was employed. Immediately following the absorption spectrum a series of spectra was recorded with successive reductions in the sector aperture. By means of a Zeiss microphotometer an accurate interpolation at a number of frequencies could be made.†

The long absorption tube consisted of seventeen sections, each 2 metres long, and 30 mm diameter, fitted together with " Normalschliffe," the whole being made by W. Heinz of Thuringer. The joints were lubricated with Apiezon " L " grease, which was found to be inappreciably attacked by hydrogen bromide. The tube was supported and accurately aligned in a teak box, the cross-section of which is shown in fig. 1. The end-pieces were made of fused quartz, and the whole was held tightly together by means of strong springs, as shown. A means of varying and controlling the temperature over a small range was also provided. The source of illumination was a quartz hydrogen discharge tube which, at 1000 volts and 1 ampere, gave intense continuous emission in the ultraviolet. With the quartz Raman spectrograph exposures could be taken in 1 minute.

PREPARATION

The hydrogen bromide was first prepared by warming KBr of A.R. quality and H_3PO_4 under reduced pressure. It was dried with P_2O_5 and freed from traces of bromine by fractional distillation. The gas so obtained was sufficiently pure for measurement in the first three ranges of Table I, but, with the long tube, bands due to traces of SO_2 appeared in the region of measurement. Direct synthesis from " atomic weight " bromine* and electrolytic hydrogen, catalysed with platinum black deposited on glass beads, gave a product containing 0·007% SO_2 as estimated by direct comparison with SO_2 bands. Bromine refluxed over BaO gave no improvement. Finally the SO_2 content was reduced to about 2 parts per million in the following manner. The bromine was left under water and then distilled through P_2O_5 in a pure oxygen stream; the hydrogen was passed through a liquid air trap and the catalyst employed was pure bright platinum gauze at a temperature of 250° C.

---

* Twyman, 'Trans. Opt. Soc.,' vol. 33, p. 9 (1931).

† Ornstein-Moll-Burger, " Objective Spektralphotometrie," (1932), p. 90 *et seq.*

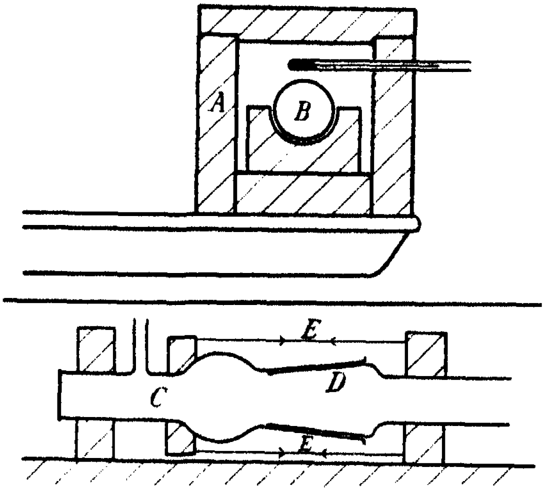


FIG. 1—A, teak box; B, absorption tube; C, quartz end-piece; D, normalschliff; E, E, springs.

This impurity constituted a limitation in the most interesting range of study. Further purification might permit an extension to extinction coefficients of 10^{-8}, but an indeterminable error due to traces of HI would become appreciable in this range.

RESULTS

The logarithms of the extinction coefficients determined, as described above, are shown plotted against the frequency in fig. 2. The methods used to obtain the respective values are indicated. The probable error in the value of α is not greater than 3% over the first three ranges and 5%

---

* Scott, ' J. Chem. Soc.,' vol. 103, p. 847 (1913).

in the long tube. The results of Dutta and of Tingley and Gerke are also shown. In the latter case the empirical assumption is made that their " absorption limit " is equivalent to a density $\left(= \log \frac{I_0}{I_T}\right)$ of 1·25. The frequencies corresponding to the energies required to dissociate HBr into H : $^2S_{\frac{1}{2}}$ and Br : $^2P_{\frac{3}{2}}$, and into H : $^2S_{\frac{1}{2}}$ and Br : $^2P_{\frac{1}{2}}$ are indicated. These values have been calculated from the thermal and spectroscopic heats given in Landolt-Bornstein's Tables (1931).

It will be seen from fig. 2 that extinction coefficients measured with pressures of about 1 atmosphere in one tube agree with those measured at low pressures in the next longer tube. The coefficient is therefore independent of pressure in this range. It was also found that an increase of temperature of about 300° C with the 1-cm tube and of 10° C with the long tube made no appreciable difference in the absorption for a fixed amount of gas.

DISCUSSION

From known constants of the hydrogen bromide molecule,*†‡ it is possible to calculate the potential energy curve for the ground state (presumably $^1\Sigma$) by means of a Morse function,§ as shown in fig. 3. From the experimental results described above and from the general equation for the probability of a change involving dipole radiation,‖ it is possible to construct the potential energy curve for the upper state which combines with the ground state in absorption. This equation may be written

$$\varepsilon = K\left[\int \psi' M \psi''\right]^2, \tag{2}$$

where ψ' and ψ'' are the molecular eigenfunctions of the upper and ground states respectively, M the dipole moment of the molecule, and ε the molecular extinction coefficient. This latter quantity is defined by the expression

$$\varepsilon = \log_{10} \frac{I_0}{I_T} \times \frac{1}{n.l.}, \tag{3}$$

where n is the number of molecules per cubic centimetre. ε has the dimensions of an area per molecule and is equal to $2 \cdot 95 \times 10^{-17}\ \alpha$. A scale of its values is shown in fig. 2. The factor K contains universal constants and the frequency, but as this latter does not change greatly over the range considered, the whole may be taken as a constant.

---

* Jevons, " Band Spectra of Diatomic Molecules " (1932), pp. 38, 276.

† Plyler and Barker, ' Phys. Rev.,' vol. 44, p. 984 (1933).

‡ Salant and Sandow, ' Phys. Rev.,' vol. 36, p. 1591 (1930); vol. 37, p. 373 (1931).

§ Morse, ' Phys. Rev.,' vol. 34, p. 57 (1929).

‖ Gibson, Rice, and Bayliss, ' Phys. Rev.,' vol. 44, p. 193 (1933).

The values of the eigenfunctions can be obtained only from a complete description of the two states, but by a series of assumptions the equation can be reduced to a form which permits the reverse calculation. According to the Franck-Condon principle, a transition between the two states occurs so rapidly that the nuclear separation is unchanged. The corresponding quantum mechanical statement given by Condon is that ε can be separated into two factors,

$$\varepsilon = \mathrm{K}\left[\int \psi'_e \mathrm{M} \psi''_e\right]^2 \left[\int_0^\infty \psi'_\mathrm{N} \psi''_\mathrm{N}\, dr\right]^2, \tag{4}$$

where ψ_e and ψ_N are the electronic and nuclear eigenfunctions respectively. The latter are functions only of the nuclear separation, r, for any particular energy level.

If for all possible transitions we confine ourselves to small changes of the nuclear separation, the electronic eigenfunctions will be unaffected and the first integral can be written as a constant,

$$\varepsilon = \mathrm{K}'\left[\int_0^\infty \psi'_\mathrm{N} \psi''_\mathrm{N}\, dr\right]. \tag{5}$$

In the following analysis the maximum displacement considered is 0·4 A, whereas the equilibrium separation, r_e, is 1·41 A.

The values of the nuclear eigenfunction for the ground state can be calculated from well-known equations. Of the possible vibrational levels we need at first consider only the zero point one, at room temperature. The eigenfunction for this state has the form

$$\psi''_\mathrm{N} = \mathrm{K}_0\, e^{-\beta(r-r_e)^2}, \tag{6}$$

neglecting the anharmonicity of the potential energy curve of the ground state, where $\mathrm{K}_0 = \sqrt[4]{\frac{2\beta}{\pi}}$ and $\beta = \frac{2\pi^2\mu}{h}\nu$, μ being the reduced mass of the molecule and ν the vibrational frequency (2560W.N.). The form of this equation is shown by the thin line in fig. 3, and has a maximum when $r = r_e$. No such simple expression can be given for the eigenfunctions of the upper state. It can be assumed, however, from the flatness of the top part of the extinction curve in fig. 2, that the slope of the upper potential energy curve is high. The eigenfunction of any state reached by the absorption of a certain frequency has therefore an appreciable value only

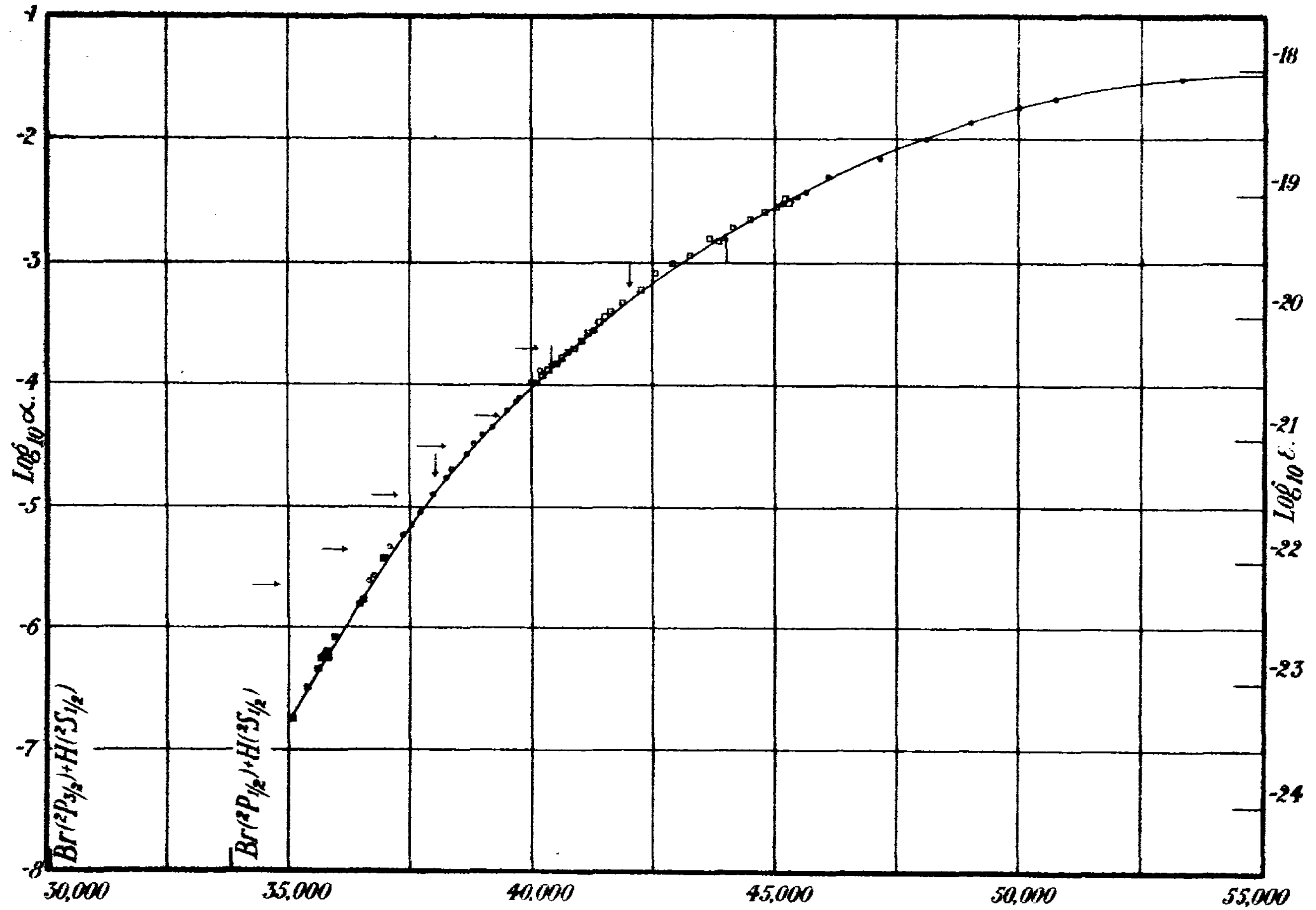


FIG. 2—The extinction coefficients of HBr. ■ 3355 cm tube; ○ 138 cm tube; □ 8 cm tube; ● 1 cm tube; → Dutta; ↓ Tingey and Gerke.

over a very small distance from the potential energy curve. The value of ψ''_N changes slowly with r for small separations and may therefore be

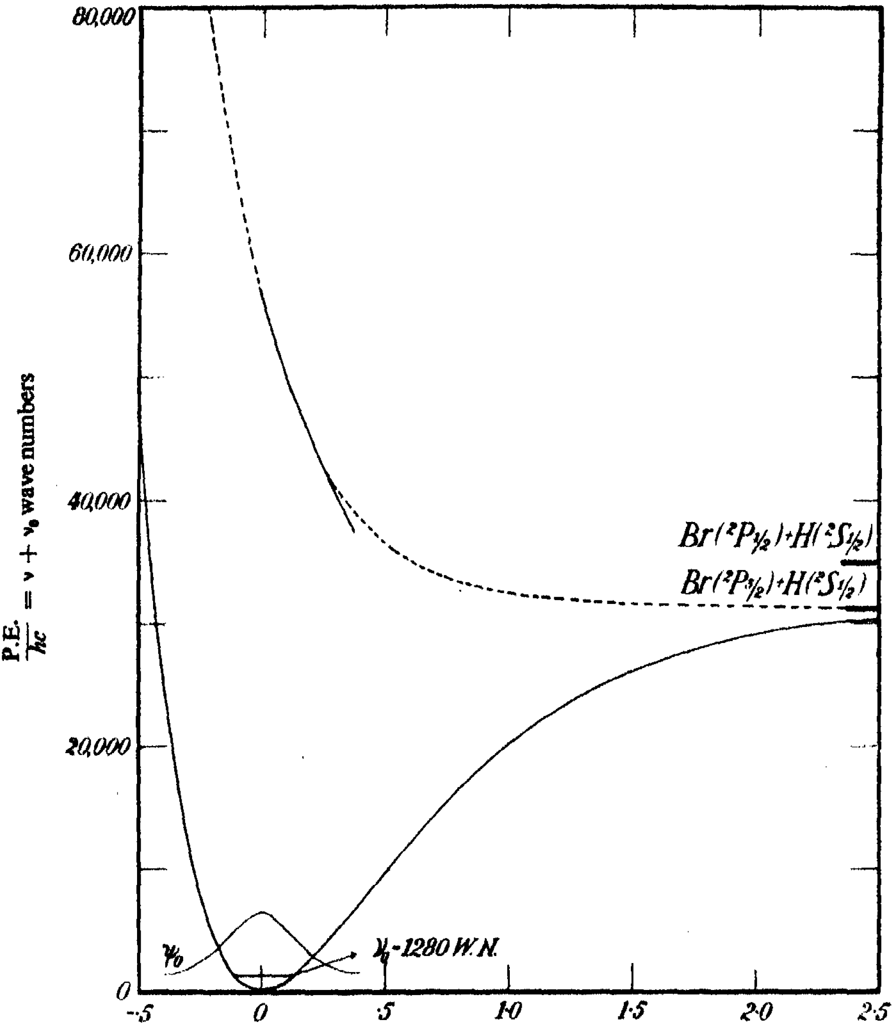


FIG. 3—Potential energy curves of HBr.

considered as constant for the range in which ψ'_N has significant values. We can therefore write for equation (5)

$$\varepsilon = \mathrm{K}'\psi''^2_N\left[\int_0^\infty \psi'_N\,dr\right]^2. \tag{7}$$

Furthermore, if the slope of the upper potential energy curve does not vary greatly over the range considered, the total area under the ψ'_N curve will be the same for all energy states, *i.e.*, the value of the integral will be a constant, and we have

$$\varepsilon = K''\psi''^2_N. \tag{8}$$

Combining this with equation (6),

$$\varepsilon = K''' e^{-2\beta(r-r_e)^2},$$

or

$$\log \varepsilon = \log K''' - 2\beta\,(r - r_e)^2.$$

When $r = r_e$, $\log \varepsilon$ is a maximum and equal to $\log K'''$. Therefore

$$\log \varepsilon_{\max} - \log \varepsilon_\nu = 2\beta\,(r - r_e)^2. \tag{9}$$

The extinction coefficient for any particular frequency is therefore a function of the distance of the point on the upper potential energy curve corresponding to this frequency change from the equilibrium position on the lower curve.

Part of the upper potential energy curve, as shown by the full line in fig. 3, has been calculated by substituting the results shown in fig. 2, in equation (9). It has been assumed that the maximum is at 56,000 wave numbers and has a value of $\log_{10} \varepsilon$ of $-18\cdot05$ cm^2. The upper portion of the curve, where the potential energy is high, is probably correct, but at the lower end the approximations used in the above derivation become less valid. It is hoped in a subsequent paper to make a more exact analysis of these results, using Gamow's function,* in the regions of "negative kinetic energy," for the ψ function of the upper state. A preliminary application of Gamow's function by a graphical method indicates that the lower potential energy end of the curve should lie at a value of $(r - r_e)$ about 0·1 A greater, as indicated by the broken curve. This curve has been drawn from the empirical equation,

$$\frac{\text{P.E.}}{hc} = \nu + \nu_0 = 31{,}330 + 25{,}950\,.\,10^{-1\cdot38\times10^8(r-r_e)}.$$

It has no minimum, but it is to be emphasized that the experimental results do not prove the absence of such a minimum, and therefore do not rule out the possibility of the upper state arising from a Br : $^2P_{\frac{1}{2}}$ atom.† It can be said, however, that the upper curve cannot approach the level corresponding to the Br : $^2P_{\frac{3}{2}}$ and H : $^2S_{\frac{1}{2}}$ atoms asymptotically. An extrapolation of the extinction coefficient to this level is therefore unjustified.

---

* Gamow "Atomic Nuclei and Radioactivity" (1931), p. 38 *et seq.*

† Results to be described in a subsequent paper indicate the absence of such a minimum with hydrogen iodide.

The omission of the first vibrational level from these considerations has been justified by calculations from the upper potential energy curve derived from equation (9). Although the extinction coefficient of molecules in the first vibrational level for lower frequencies is found to be considerably greater than that of molecules in the zero level, at room temperature the number of molecules in the first vibrational level is only 10^{-5} of the total number of molecules. The total extinction due to molecules in this level is found to be less than one-tenth of the observed extinction. In support of this, we know that a rise of temperature of *10° C increases the number of molecules in this level by 50%, and yet no* increase in extinction is observed. Similarly the effect of rotation is small and can be neglected.

A discussion of the possible potential energy curves arising out of the interaction of a hydrogen and a bromine atom is beyond the scope of this paper. It is sufficient to say that there are probably four such curves for the Br : $^2P_{\frac{3}{2}}$ atom, corresponding to two singlet and two triplet states. As the extinction coefficient has normal values, it is to be presumed that the upper curve, as well as the lower, corresponds to a singlet state. Transitions to the triplet states may possibly be contributing in the region of very low extinctions where the values are 10^{-6} of that at the maximum. Spectrophotographs taken with the long tube and with plates sensitive up to 11,000 A have revealed no further absorption, and therefore the probability of a transition to such triplet states must be less than 10^{-6} of that to the singlet state.

The authors are deeply indebted to Dr. E. Teller for his valuable assistance in the theoretical interpretation of the above results.

SUMMARY

The extinction coefficients of hydrogen bromide have been measured over a wide range of pressures and tube lengths. The approximate course of the upper potential energy curve has been calculated from the eigenfunction of the ground state and the observed extinctions. The results favour the dissociation into normal atoms, and indicate the difficulties of extrapolation of absorption thresholds in the determination of heats of dissociation.

The Effect of Thermal Agitation on Atomic Arrangement in Alloys—III

By E. J. WILLIAMS, Manchester University

(*Communicated by W. L. Bragg, F.R.S.—Received June* 11, 1935)

1—INTRODUCTION

The various theoretical treatments of the atomic arrangement in alloys which have recently been given by different workers have been discussed in a paper by Professor W. L. Bragg and the writer in the last number of the 'Proceedings.'[*] The purpose of this paper is in the first place to discuss further the relation between the free energy of the alloy and the quantities occurring in the earlier treatment of the problem by these authors[†] and to give the detailed calculations of the equilibrium equations for a general case, first, from the standpoint of the Boltzmann distribution formula, and, secondly, from the principle of minimum free energy.

The treatment recently given by Bethe differs essentially from previous treatments[‡] in that he assumes the energy associated with the atomic arrangement to be uniquely determined by the order of neighbours (σ) and not by the superlattice order S. By comparing the dependence of the entropy of the atomic arrangement on σ with its dependence on S the quantitative relation of the two types of treatment may readily be seen. This is done in the present paper. It is further shown that an approximate treatment of the problem, assuming interaction only between nearest neighbours, may be given by combining the formula of Bragg and Williams for superlattice order with an approximate formula for " liquid " order.

In view of the appreciable deviation in certain respects of the experimental results from the requirements of existing calculations, a review of the assumptions involved in these calculations is given at the end of the paper and the possible modifications discussed.

The paper is arranged as follows:

§ 2—Derivation of equilibrium equation for a general case, first, in B-W[§] type of treatment and, secondly, from principle of minimum free energy.

§ 3—Entropy and energy in B-W treatment.

---

[*] 'Proc. Roy. Soc.,' A, vol. 151, p. 540 (1935), referred to as Part II.

[†] 'Proc. Roy. Soc.,' A, vol. 145, p. 699 (1934). We shall refer to this paper as Part I.

[‡] Detailed references are given in Part II.

[§] Denotes Bragg and Williams.

§ 4—Application of B-W treatment assuming Bethe's interaction energies between neighbouring atoms.

§ 5—Entropy and energy in Bethe's treatment.

§ 6—Approximate treatment assuming interaction only between nearest neighbours allowing for local order.

§ 7—General discussion.

2—DERIVATION OF EQUILIBRIUM EQUATION FOR A GENERAL CASE IN B-W TYPE OF TREATMENT AND FROM PRINCIPLE OF MINIMUM FREE ENERGY

The problem under consideration is the effect of temperature on the arrangement of the different kinds of atoms in an alloy when the structure of the lattice points is given. In a given mass of an alloy, let there be N atoms altogether, and therefore N lattice points (or sites). In actual problems the nature of the atoms occupying a certain simple fraction of these N sites may be practically unaffected by thermal agitation, so that in general we are concerned with the occupation of a certain number $n, < \mathrm{N}$, of the sites. Let an be the number of A atoms to be distributed amongst these n sites so that the number of B atoms is $(1 - a)\, n$. For complete disorder the probability of any one of the n sites being occupied by an A atom is a. In all existing treatments, except that of Bethe, the only kind of order which is recognized is that in which similar atoms form on their own a superlattice which repeats itself throughout the crystal, the A atoms forming one superlattice, the B atoms another, the two lattices meshing with each other. We shall refer to this type of order as superlattice order, or, as Bethe calls it, " long-distance " order.

Let the number of lattice points in the A superlattice be rn (α-positions). The number in the B superlattice is then $(1 - r)\, n$ (β-positions). If $a = r$ we have the simple atomic ratio considered in Part I. $a \neq r$ corresponds to the more general case of departure from simple atomic ratio. In accordance with the definition of order for $r = a$, given in Part I, the degree of superlattice order S in the general case may be written

$$\mathrm{S} = (p - a)/(q - a); \quad q = 1 \text{ if } a > r, \quad q = a/r \text{ if } a < r, \tag{1}$$

p denotes the probability that an α-position is occupied by an A atom, q the maximum value of p, while a represents the value of p for complete disorder.

The simplifying assumption which characterizes the calculations of Bragg and Williams is that all the α-positions are identical and all the β-positions identical, so that the interchange of an A atom in an α-position with a B atom in a β-position, means a unique increase V in internal energy.* This makes the internal energy U associated with the atomic arrangement a unique function of the superlattice order S, the relation between V and U being

$$\mathrm{V}(\mathrm{S}) = -(rn\overline{q-a})^{-1}\,\partial \mathrm{U}(\mathrm{S})/\partial \mathrm{S}. \tag{2}†$$

At a given temperature T the statistical distribution of A and B atoms between α- and β-positions is determined by the value of V. In Bragg and Williams's calculations, this distribution is obtained by applying Boltzmann's distribution formula. We have

$$\frac{\text{number of A atoms in } \alpha\text{-positions}}{\text{number of A atoms in } \beta\text{-positions}} =$$

$$\frac{\text{number of } \alpha\text{-positions not occupied by A atoms}}{\text{number of } \beta\text{-positions not occupied by A atoms}}\, e^{\mathrm{V}/k\mathrm{T}}$$

$$\frac{rnp}{n(a-rp)} = \frac{rn(1-p)}{n(1-r-a+rp)}\, e^{\mathrm{V}/k\mathrm{T}}$$

$$\mathrm{V}/k\mathrm{T} = \log\{p(1-r-a+rp)/(1-p)(a-rp)\} = f_1(\mathrm{S}). \tag{3}$$

This determines the degree of order allowed by statistical mechanics for a given value of the ordering energy V.

We shall now consider the derivation of this formula from the principle of minimum free energy. We have

$$\mathrm{F}(\mathrm{S}) = \mathrm{U}(\mathrm{S}) - \mathrm{T}\phi(\mathrm{S}). \tag{4}$$

F denotes the free energy and ϕ the entropy associated with the atomic arrangement. To determine ϕ (S) we use the relation between the entropy of a system and the number of different " microscopic " states of the system which is possible under the given conditions. From this relation if the possibility of interchanging unlike atoms in a given mass of an alloy increases the number of possible microscopic states by a factor W, then it increases the entropy of the alloy by k log W where k is the atomic gas-constant. The dependence of the entropy of the alloy on the degree of order may therefore be obtained by evaluating W for a given value of S.

---

* Regarding this assumption of the uniqueness of V, it was recognized by Bragg and Williams in their paper that actually the atomic arrangement in any small region would fluctuate and be only described on the average by the superlattice order S. In view of the corresponding fluctuations in V that this might give rise to, it was pointed out that the V used in the calculations was to be regarded as the *effective average* value of V corresponding to the degree of order S. The effect of these fluctuations is one of the main things which the recent calculations of Bethe take into account.

† The notation of partial differentials is used here and throughout the paper, though under the special assumptions made this is not everywhere necessary.

When the degree of order is S, $= (p - a)/(q - a)$, there are *rnp* A atoms in positions of order α, and $rn(1 - p)$ B atoms. The total number, m_1, of complexions which the occupation of the α-positions can have through the possibility of interchanging unlike atoms is then the number of combinations of *rn* things taken *rnp* at a time,* *i.e.*,

$$m_1 = {}^{rn}C_{rnp} = (rn)\,!/(rn - rnp)\,!\,(rnp)\,!$$

Similarly the number of complexions which the occupation of the β-positions can have is

$$m_2 = {}^{n-rn}C_{an-rnp} = (n - rn)\,!/(n - rn - an + rnp)\,!\,(an - rnp)\,!$$

The value of W is evidently the product of m_1 and m_2 and the dependence of entropy on order is given by

$$\phi(p) = k \log W = k \log (m_1 m_2).$$

Applying Stirling's theorem, viz., that for large N, $N\,! = (N/\varepsilon)^N$, this gives

$$\begin{aligned}\phi(p) = C - nk\,\{&r(1 - p)\log(1 - p) + rp\log p \\ &+ (1 - r - a + rp)\log(1 - r - a + rp) + (a - rp)\log(a - rp)\} \quad (5)\end{aligned}$$

where C is a constant independent of *p*.

$$\begin{aligned}\partial\phi/\partial S &= (q - a)\,\partial\phi/\partial p \\ &= -(q - a)\,rnk\log\{p(1 - r - a + rp)/(1 - p)(a - rp)\}. \quad (6)\end{aligned}$$

At equilibrium the free energy is a minimum, *i.e.*,

$$\partial\phi/\partial S = T^{-1}\partial U/\partial S.$$

From (2) $\partial U/\partial S = -(q - a)\,rnV$, so that the equilibrium degree of order for a given value of V is given by

$$V/T = k\log\{p(1 - r - a + rp)/(1 - p)(a - rp)\}. \quad (7)$$

This is identical with the equilibrium equation (3).

3—REPRESENTATION OF ENTROPY AND ENERGY IN CALCULATIONS OF BRAGG AND WILLIAMS

In these calculations there is another relation between V and S in addition to the statistical relation (3). This second relation expresses the effect of the degree of order on the ordering forces, and it was shown that it must be approximately of the form $V = V_0 S$, but for generality we may write it here:

---

* Being the number of ways in which the *rnp* A atoms can be distributed in the *rn* positions—for when the A atoms are all fixed there is no choice in the distribution of the B atoms.

$$V/kT = f_2(S). \tag{8}$$

To obtain the actual equilibrium state this value of V must be used in (3), so that the equilibrium value of S is that which satisfies

$$f_2(S) = f_1(S). \tag{9}$$

In their actual calculations Bragg and Williams plot the two functions f_1 and f_2 against S, the equilibrium state being then determined by the intersections of the two curves. We wish here to point out the following simple relations between these curves and the entropy and energy of the alloy. From (2) and (8)

$$f_2(S) = -(rnk\overline{q-a}T)^{-1}\partial U/\partial S = -AT^{-1}\partial U/\partial S. \tag{10}$$

From (3) and (6),

$$f_1(S) = -(rnk\overline{q-a})^{-1}\partial\phi/\partial S = -A\partial\phi/\partial S. \tag{11}$$

The two curves plotted by Bragg and Williams thus represent the rate of change with order of the entropy and of (energy ÷ temperature) respectively. The difference between them therefore represents the rate of change of free energy F with order.

Actually

$$(\partial F/\partial S)_T = \partial U/\partial S - T\partial\phi/\partial S = -A^{-1}T(f_2(S) - f_1(S)). \tag{12}$$

An intersection of the two curves thus means a stationary value of the free energy and therefore an equilibrium state. The equilibrium is unstable if the stationary value is a maximum, *i.e.*, $\partial^2 F/\partial S^2 < 0$ and therefore $\partial f_2/\partial S > \partial f_1/\partial S$. It is stable, or metastable, if $\partial^2 F/\partial S^2 > 0$, *i.e.*, $\partial f_2/\partial S < \partial f_1/\partial S$. These criteria are the same as those previously derived by Bragg and Williams.

The above interpretation of the curves $f_1(S)$ and $f_2(S)$ enables us to go farther than before in that we can distinguish the permanently stable state from a metastable state. Both give a minimum in the free energy. The permanently stable state is that which gives the lower minimum. In order to distinguish the latter we use (12), according to which the difference in free energy between two states of order S_1 and S_2, at a given temperature, is

$$F_2 - F_1 = \int_{S_1}^{S_2} (\partial F/\partial S)dS = AT\int_{S_1}^{S_2} (f_1(S) - f_2(S))\,dS. \tag{13}$$

It is therefore represented by the area bounded by the two curves $f_1(S)$ and $f_2(S)$ and the lines $S = S_1$, and $S = S_2$. In fig. 1 it is the area HIJK. The total area GIJL under the curve f_2 (S) represents the difference in (internal energy ÷ T) between the two states, and the area GHKL, under $f_1(S)$, the difference in entropy between the two states.

If now the curves f_1 and f_2 have two intersections which give a minimum in free energy, as at O and B in fig. 1, the one at B is the permanently stable one, or not, according to whether the area under the curve f_2 (projected on S axis) between O and B is greater or less than the corresponding area under f_1. This means that B is the stable state if the area of the loop AHKBA is greater than that of the loop OFAO, and vice versa. If the two loops have the *same* area, both states have the same free energy, and at the temperature for which this is so both states can coexist in permanent equilibrium. This is therefore the true critical temperature T at which a reversible transition from the state of complete disorder at O to partial order can take place. It is shown in Part I that for an alloy with $a = r = V\frac{1}{4}$, the curves $f_2(S)$ and $f_1(S)$ are of the form giving triple intersection. By drawing the curves for different temperatures and applying the above criterion we find that the true critical temperature is equal to $0 \cdot 205$ (V/k), being very close to the temperature at which a metastable ordered state is first possible, viz., $0 \cdot 207$ (V/k).

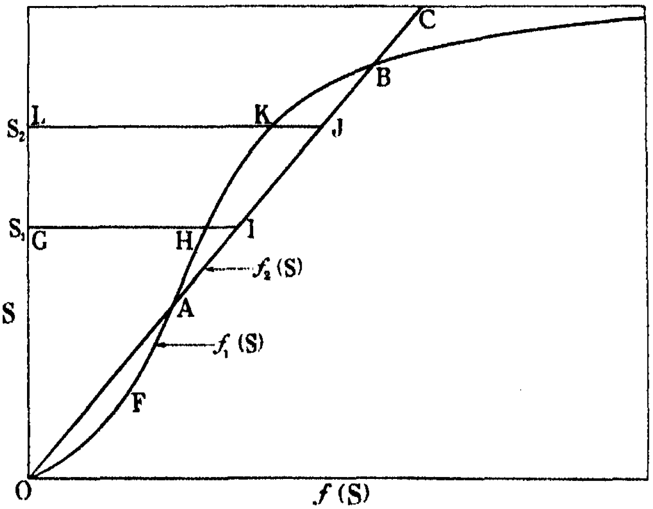


FIG. 1.

The information about the free energy of the alloy obtained from the curves f_1 and f_2 is of course exactly the same as that given by a direct analysis in terms of free energy as in § 2. The above discussion has been given to show how the various cases considered by Bragg and Williams in terms of f_1 and f_2 may be translated into terms of free energy. This translation has been carried out in Part II, in accordance with these rules, for the triple intersection of the curves f_1 and f_2, and also for no triple intersection, the purpose being to illustrate further the nature of the order-disorder transformation in alloys and its relation to a phase change.

4—APPLICATION OF B-W TYPE OF TREATMENT, ASSUMING BETHE'S INTERACTION ENERGIES BETWEEN NEIGHBOURING ATOMS

The calculation of superlattice order in Part I is based on the relation $V = V_0S$, where V is the energy required to interchange two unlike atoms initially in positions of order, when the degree of superlattice order is S. V is connected with the energy of order, U(S), according to (2), viz., $V(S) = -(nr\, q\overline{-a})^{-1}\partial U(S)/\partial S$. Now if we assume with Bethe (*see* § 3, Part II) that the energy associated with the atomic arrangement is due to the mutual energies of atoms which are nearest neighbours—V_{aa} for an AA pair, V_{bb} for a BB pair, and V_{ab} for an AB pair—the energy U(S) may be obtained by calculating the order of neighbours, σ, *i.e.*, the numbers N_{ab}, N_{aa}, and N_{bb}, of the different kinds of pairs which exist in the crystal when the degree of order is S. Actually, since we are only concerned with the change of energy with S, it is sufficient to calculate the dependence only of N_{ab} on S, since the changes in N_{aa} and N_{bb} are always opposite to and equal to one-half the change in N_{ab}.

A given value of the superlattice order S determines the partition of the atoms between positions of order and disorder, but it leaves unrestricted their detailed distributions amongst these positions. As a result a given value of S does not correspond to a unique value of σ. It is assumed in the calculations of Bragg and Williams, and of Borelius and Dehlinger, that these detailed distributions are random, *i.e.*, that the A atoms in their positions of order α, for instance, are randomly distributed amongst those positions. The number N_{ab}, being very large, may then be taken equal to its average value for such distributions, and it may be shown that this average value is

$$N_{ab} = nZa(1-a) + nZr^2(1-r)^{-2}(q-a)^2S^2. \qquad (14)$$

The notation of § 2 is used, *i.e.*, *rn* denotes the number of positions of order (α) for A atoms and *an* the number of A atoms present. It is assumed that every lattice position has Z nearest neighbouring positions and that for

$r < 1/2$ each α-position is surrounded only by β-positions, and that all β-positions are surrounded by the same number of α-positions.

The order of neighbours, σ, may be defined for the general case as

$$\sigma = (f - f_0)/(f_1 - f_0), \qquad (15)$$

where f is the fraction of pairs which are unlike, f_0 is the value of f for complete disorder, and f_1 is the maximum possible value of f. Substituting for these their values given by (14), we obtain the simple relation

$$\sigma = S^2. \qquad (16)$$

Assuming Bethe's interaction energies, we have

$$\begin{aligned} U(\sigma) - U(\sigma = 0) &= \{n_{ab}(\sigma) - n_{ab}(\sigma = 0)\} \tfrac{1}{2}(2V_{ab} - V_{aa} - V_{bb}) \\ &= -nZr^2(1 - r)^{-2}(q - a)^2 V_B \sigma, \qquad (17\text{A}) \end{aligned}$$

so that, introducing the same energies into the B-W type of treatment, we have, by virtue of (16),

$$U(S) - U(S = O) = -nZr^2(1 - r)^{-2}(q - a)^2 V_B S^2. \qquad (17\text{B})$$

It follows that

$$V = -(rn\overline{q - a})^{-1} \partial U/\partial S = 2r(1 - r)^{-2}(q - a) ZV_B S. \qquad (18\text{A})$$

Since the coefficient of S on the *r.h.s.* is independent of S this is precisely the same as the relation between V and S used by Bragg and Williams, viz., $V = V_0 S$, provided we take

$$V_0 = 2r(1 - r)^{-2}(q - a) Z V_B. \qquad (18\text{B})$$

This formula correlates the energy V_0 with the energy V_B used by Bethe, and shows how it depends on the value of r and on the composition a. The dependence of the equilibrium state on these quantities may therefore be obtained, to the approximation of the B-W treatment, by combining (18) with the equilibrium equation (3). It is found, for instance, that the general expression for the critical temperature T'_c, at which complete disorder becomes unstable—which is given by $\partial f_1(S)/\partial S = \partial f_2(S)/\partial S$ at $S = 0$ — is

$$T'_c = (2ZV_B/k)\, a(1 - a)\, r(1 - r)^{-1}. \qquad (19)^{*}$$

Though it is thus possible to use Bethe's interaction energies in the B-W method, the results are, of course, not those which the Bethe type of treat-

---

[*] Formulae (15) and (17) in § 5 of Part II are special cases of this formula for $r = 1/2$ and $r = 1/4$ respectively.

ment would give. On Bethe's assumption the energy of the atomic arrangement is a function of the order of neighbours σ, and is not a unique function of the superlattice order S. The error in the foregoing discussion comes in the simplifying assumption that the atoms in positions of order and disorder are randomly distributed in those positions. Actually, if we consider, for instance, the β-positions surrounding a given α-position, the fraction of these occupied by B atoms will be greater than the average fraction for all β-positions throughout the crystal if the central α-position is occupied by an A atom. It will be less than the average if the central position is occupied by a B atom. This local order is due to the greater attraction between unlike neighbours than between like neighbours. It exists over and above the superlattice order and makes the number N_{ab} of unlike pairs, for a *given* S, greater than that given by (14). This effect of local association of unlike atoms is allowed for in Bethe's calculations.

It is interesting to consider two limiting conditions for which the effect of local association becomes vanishingly small. We have seen that the effect invalidates the calculation of N_{ab} given by (14), because an A atom in an α-position, for instance, tends to make the fraction of B atoms in the surrounding β-positions greater than the average value for β-positions throughout the crystal. Now for an atom in any one of these neighbouring β-positions the central atom A is only one neighbour in Z, so that the influence it has on the average nature of its neighbours becomes negligible for large Z. Under these conditions the distribution of atoms in the positions surrounding an α-position is accordingly independent of whether the α-position concerned is occupied by an A or a B atom. The above derivation of the formula (14) for N_{ab} is then valid. It must, however, be remembered that (14) refers to the average value of N_{ab}, so that it is the average value of V that is really represented by (18). However, if aZ is large, as well as Z, the value of V for individual transitions from positions of order to disorder will be the same as the average value, because the relative fluctuations in the nature of any given position vanish when the number of different kinds of nearest neighbours is infinitely large. It follows that for large Z, or strictly speaking large aZ, the necessary conditions for the applicability of the method of Bragg and Williams to the problem considered by Bethe are satisfied, and the method provides an exact solution of it. Bethe has made a comparison of the actual results of the two treatments for $a = r = 1/2$, and in accordance with this general conclusion finds them to be identical for $Z = \infty$. In actual alloys the value of Z are of the order 6 to 12. The discussion in § 6 provides a general method of estimating the extent to which the B-W solution diverges from the exact solution in such cases.

Bethe's treatment also merges into that of Bragg and Williams when the degree of long-distance order is high. For if the number of atoms in positions of disorder is small they will be sufficiently dispersed to make V equal to V_0 nearly everywhere, and this is the state of affairs assumed by Bragg and Williams when S is close to unity.

5—Entropy and Energy in Bethe's Treatment

Though Bethe in his calculations does not consider the entropy explicitly, his final results for the equilibrium degree of order may readily be translated back into terms of entropy, and it is instructive to consider the results.

The energy of the atomic arrangement in Bethe's treatment is determined by the degree of order of nearest neighbours σ, defined for the general case by (15). Writing s^2 for σ, we have for the general case (*see* 17A),

$$U(s) - U(o) = -nZr^2(1-r)^{-2}(q-a)^2 V_B s^2. \tag{20}$$

For equilibrium $\partial\phi(s)/\partial s = T^{-1}\partial u(s)/\partial s$ so that

$$\partial\phi(s)/\partial s = -2nZ(V_B/T)\,r^2(1-r)^{-2}(q-a)^2 s \tag{21}$$

$$= -2nZ(V_B/T)\,r^2 s, \quad \text{if} \quad a = r. \tag{22}$$

For the special case considered by Bethe, viz., $a = r = 1/2$ this gives

$$\partial\phi(s)/\partial s = -\tfrac{1}{2}nZ(V_B/T)\,s. \tag{23}$$

The value of $\partial\phi(s)/\partial s$ for a given value s' of s according to Bethe's calculations may therefore be obtained from his results for the variation of the equilibrium value of s with temperature, being equal to $\tfrac{1}{2}nZ(V_B/T^1)\,s'$, where T' is the temperature for which s' is the equilibrium value of s. The variation of $\partial\phi(s)/\partial s$ with s obtained in this way from Bethe's results is represented in fig. 2 by the curve OAB. The curve actually represents $(nk)^{-1}\partial\phi/\partial s$, *i.e.*, the rate of change of entropy with s per atom of the alloy, expressed in terms of k as unit.

The curve OCD represents $(nk)^{-1}\partial\phi/\partial s$, according to the approximate calculation of Bragg and Williams.* The two curves approach each other as s approaches unity, in accordance with the conclusions arrived at in the previous section. They differ widely for small s. For the initial part OLA of Bethe's curve there is no superlattice or long-distance order. In this region of small s there are fewer microscopic states of long-distance order which give the requisite number of unlike pairs than there are of no long-distance order. For $a = r = 1/2$, it is possible to have *all* neighbouring pairs unlike by having complete long-distance order. When the number of unlike pairs is only a little less than this maximum there will still be long-distance order, because under such conditions the number of microscopic states for which the wrong couplings are distributed so as to divide the crystal into equal volumes which are " out of step "—thus destroying long-distance order—is negligible compared with the number of states in which the wrong couplings are randomly distributed. Thus, while there is no long-distance order for small s there is of necessity such order when s is close to unity. There must therefore be some critical value of s at which long-distance order becomes a normal property. This is the point A in the curve OAB. For any given value of s below A long-distance order, however small, is of the nature of an extra restriction which greatly reduces the possible number of microscopic states. Above A, on the other hand, the number of microscopic states without long-distance order is small compared with the number with such order—in fact, negligibly small, so that the normal state is one with long-distance order. At A, the curve changes its slope and this gives rise to the phenomenon of the critical temperature. On the high-order side of A, the slope is practically the same as the initial part of the approximate curve OCD.

---

* Apart from a numerical constant $(nk)^{-1}\,\partial\phi/\partial s$ is equivalent to the quantity $f_1(s)$, plotted by Bragg and Williams (*see* § 3, equation (11)). $(nkT)^{-1}\,\partial u/\partial s$ is related in the same way to the other function $f_2(s)$ in the B-W treatment (10). In this and the following sections we have considered it better to refer directly to the quantities $(nk)^{-1}\,\partial\phi/\partial s$ and $(nkT)^{-1}\,\partial u/\partial s$ than use the corresponding functions $f_1(s)$ and $f_2(s)$.

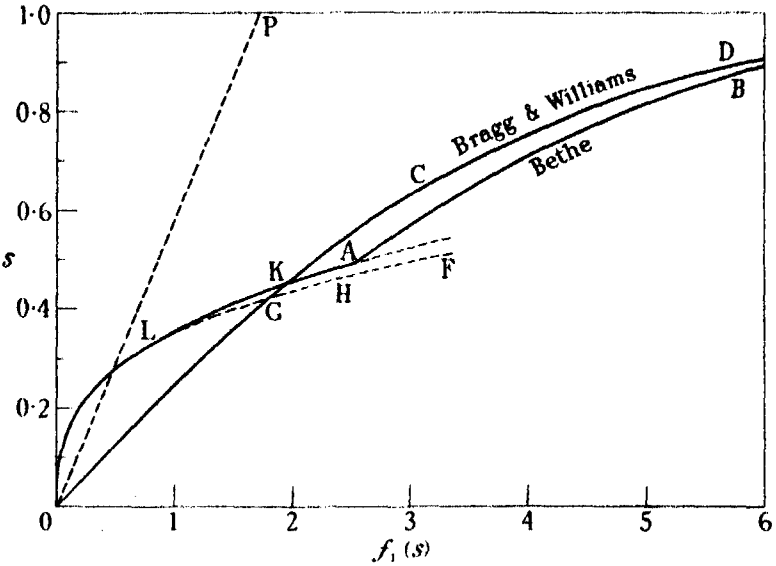


FIG. 2.

It is of interest to consider the total area between the s axis and the curves OAB and OCD. This area represents the difference between the entropy ϕ_0, of the completely disordered state and the entropy, ϕ_1, of the ordered state. This is correctly given by the calculations of Bragg and Williams.* The area under OAB, which represents Bethe's results, should therefore be equal to that under OCD. In other words, the area of the loop OLKO should be equal to the area KDBAK. It will be seen that this is very nearly so. This test of Bethe's calculations is, in fact, exactly equivalent to the test which Bethe applies in the form

$$\int_{T=0}^{\infty} T^{-1}\, dU\,(T) = \phi_0 - \phi_1.$$

We see, however, that since it is exactly satisfied by the simple treatment of Bragg and Williams it is not a very stringent test of the higher approximations of Bethe's theory.

The equilibrium state for any temperature is given by the intersection of the curve representing $(nk)^{-1}\partial\phi/\partial s$ and the curve $(nkT)^{-1}\partial U/\partial s$. U (s), which represents the internal energy, is given by (20), and for $a = r = 1/2$, $(nkT)^{-1}\partial U/\partial s = (\frac{1}{2}ZV_B/kT)\,s, = (3V_B/kT)\,s$ for $Z = 6$. This is represented by the broken line OP in fig. 2. The fact that OAB has a vertical tangent at O means that local order does not vanish for finite temperatures. The critical temperature in Bethe's theory is that which makes OP pass through A. Above this temperature there is only local order—a state of the alloy which Bethe refers to as liquid order.

6—APPROXIMATE TREATMENT, ASSUMING INTERACTION ONLY BETWEEN NEIGHBOURING ATOMS ALLOWING FOR LOCAL ORDER

If we assume interaction only between nearest neighbours, as in Bethe's treatment, the problem of determining the equilibrium state is mainly the problem of determining the dependence of entropy on the number of unlike nearest neighbours, *i.e.*, on the parameter σ. In the preceding section we

---

* From (6) we find for $a = r = 1/2$,

$$\phi_0 - \phi_1 = \tfrac{1}{2}nk\Big[(1+S)\log(1+S) + (1-S)\log(1-S)\Big]_0^1 = nk\log 2.$$

Now $\phi_0 - \phi_1 = k\log W_0/W_1$, where W_0 = number of microscopic states for complete disorder, and W_1 = number of states for complete order.

$$W_1 = {}^{n}C_{\frac{1}{2}n} = n!/(\tfrac{1}{2}n)! \qquad W_0 = 2.$$

Hence

$$\phi_0 - \phi_1 = k\,(\log W_0 - \log W_1) \approx k\,(n\log 2 - \log 2) \approx nk\log 2$$

in agreement with (6).

have given the curve for $\partial\phi/\partial s$, $s = \sqrt{\sigma}$, for an alloy with equal numbers of unlike atoms ($a = r = 1/2$) and $Z = 6$, according to Bethe's detailed results for this special case. We shall now consider an approximation to the detailed method used by Bethe and discuss other cases than that of $r = 1/2$, $Z = 6$. We shall assume that $a = r$, to simplify the formulae.

We have already seen that for large s (or σ) the calculations of Bragg and Williams provide a good approximation to the value of $\partial\phi/\partial s$ for superlattice order. For simple composition, $a = r$, their calculation gives from (11), (3), and (1),

$$-(nk)^{-1}\,\partial\phi/\partial s = r(1-r)f_1(S) =$$

$$r(1-r)\log\left\{\left(1+\frac{r}{1-r}s\right)\left(1+\frac{1-r}{r}s\right)(1-s)^{-2}\right\}. \quad (24)$$

An approximation to $\partial\phi/\partial s$, when there is no superlattice order, may be obtained from considerations of liquid order. When the order of neighbours is σ the concentration of AB, AA, and BB pairs are respectively proportional to $(r\,\overline{1-r}+r^2\sigma)$, $(r^2-r^2\sigma)$ and $(\overline{1-r}^2-r^2\sigma)$, and if there is no superlattice order an approximate representation of the equilibrium state may be obtained by applying mass action considerations to the process $AA + BB \rightarrow 2AB + \varepsilon$, where $\varepsilon = V_{aa} + V_{bb} - 2V_{ab} = 2V_B$. We then have

$$(r\,\overline{1-r}+r^2\sigma)^2\,(r^2-r^2\sigma)^{-1}\,(\overline{1-r}^2-r^2\sigma)^{-1} = e^{2V_B/kT}. \quad (25)$$

For $r = 1/2$ this gives $(1+\sigma)/(1-\sigma) = e^{V_B/kT}$ in agreement with Bethe's formula for liquid order for small σ. In terms of s, $= \sqrt{\sigma}$, (25) together with (22) gives

$$-(nk)^{-1}\,\partial\phi/\partial s = 2Zr^2sV_B/kT =$$

$$r^2Zs\log\left\{\left(1+\frac{r}{1-r}s^2\right)\left(1-\left(\frac{r}{1-r}\right)^2s^2\right)(1-s^2)^{-1}\right\}. \quad (26)$$

The curve OLF in fig. 2 represents, for $r = 1/2$ and $Z = 6$, the value of $(nk)^{-1}\,\partial\phi/\partial s$ according to these calculations. In the region of s, where there is no superlattice order, it will be seen to be a good approximation to the curve OLA which represents Bethe's detailed calculations. Beyond this region the curve OCD is a good approximation, and a combination of the two curves, represented by the composite OLGGCD, gives quantitative results quite close to those given by Bethe's curve OAB. In fact, since we know that the total area under the correct curve is the same as that under OCD, we should follow the curve OLF beyond G to some point such as H, chosen so that when the curve is brought back again to the curve OCD at large s ($\sim 0 \cdot 95$), the area HGCDH is equal to that of the loop OLG. By doing this we obtain results which differ negligibly from those of Bethe. It is interesting to apply this approximation to other cases than that of $r = 1/2$, $Z = 6$. In doing so it is much simpler to use σ itself as variable than $s = \sqrt{\sigma}$, as in fig. 2. In terms of σ the expression (26) for liquid order becomes

$$-(nk)^{-1}\partial\phi/\partial\sigma = \tfrac{1}{2}r^2 Z \log\left\{\left(1 + \frac{r}{1-r}\,\sigma\right)\left(1 + \frac{1-r}{r}\,\sigma\right)^{-1}(1-\sigma)^{-1}\right\}, \quad (27)$$

and (24) for superlattice order gives

$$-(nk)^{-1}\partial\phi/\partial\sigma = r(1-r)(2\sqrt{\sigma})^{-1}$$

$$\log\left\{\left(1 + \frac{r}{1-r}\sqrt{\sigma}\right)\left(1 + \frac{1-r}{r}\sqrt{\sigma}\right)(1-\sqrt{\sigma})^{-2}\right\}. \quad (28)$$

Also

$$-(nkT)^{-1}\partial U/\partial\sigma = r^2 Z V_B/kT. \quad (29)$$

It will be noticed that this is independent of σ, so that it is represented by a straight line parallel to the σ axis, and at a distance $r^2 Z V_B/kT$ from it.

OC_6D in fig. 3 represents in terms of σ the value of $(nk)^{-1}\partial\phi/\partial\sigma$ for $a = r = 1/2$, $Z = 6$, which is represented in terms of s by OAB in fig. 2. AGD in fig. 3 represents $(nk)^{-1}\partial\phi/\partial\sigma$ according to the treatment of Bragg and Williams. Since this curve gives the correct total change in entropy between $\sigma = 0$ and $\sigma = 1$, it follows, as already explained in connexion with fig. 2, that the critical point C_6 must satisfy the area condition $G_6DC_6 = OG_6A$. The critical temperature T_{C6} is that which makes PQ pass through C_6, *i.e.*, $\tfrac{1}{4}6V_B/kT_{C6} = ON_6$. According to the figure $ON_6 = 0 \cdot 66$, so that $kT_{C6} = 2 \cdot 27\, V_B$, as compared with $2 \cdot 35 V_B$ according to Bethe's more detailed calculations.

The change from $Z = 6$ to $Z = 8$ is simply a change in the slope of the line OF representing local order. For the small values of σ concerned the formula (27) for liquid order reduces to $\tfrac{1}{4}Z\sigma$, so that OF is approximately a straight line with a slope of $\tfrac{1}{4}Z$. OF_8, which represents the case of $Z = 8$, is therefore closer to the axis of abscissa than OF_6 ($Z = 6$) in the ratio $6:8$. This makes the area of the loop OG_8A less than that of the loop OG_6A. Since $G_8DC_8 = OG_8A$, and $G_6DC_6 = OG_6A$, this means that the critical point C_8 is closer than C_6 to the critical point A of the B-W treatment. Thus, as Z increases, C moves towards A along the line BA, and must ultimately coincide with A for $Z = \infty$. The solution for $Z = \infty$ is thus identical with that of the simple treatment. This result is already proved in § 4. The present considerations, however, enable us to see the nature and magnitude of the failure of the simple treatment when Z is finite.

A quantity of special interest is the ratio of the energy of disorder at the critical temperature to kT_C. The former is proportional to $1 - \sigma_C$, where σ_C is the value of σ at the critical point. kT_C is inversely proportional to ON. Assuming AB to be, to a first approximation, a straight line, we find that

$$\{U(T_C) - U(0)\}/kT_C, = \rho \text{ say}, = 0\cdot 50 - a/Z,$$

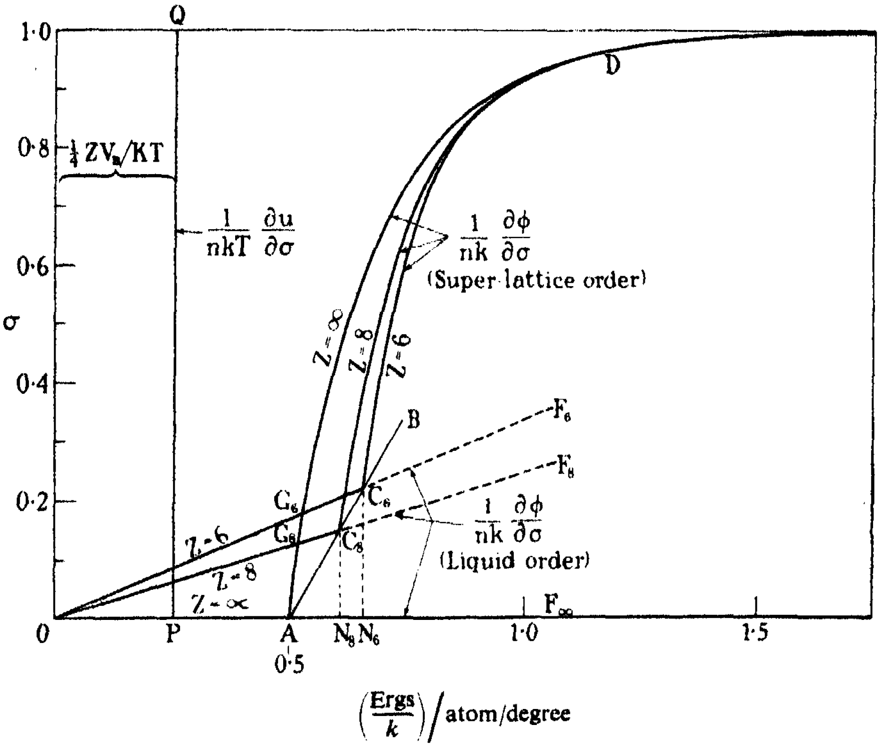


FIG. 3.

where a is a constant. If we assume Bethe's result for $Z = 6$, viz., $\rho = 0\cdot 48$, we have $a = 0\cdot 12$, and therefore for any Z

$$\rho = 0\cdot 50 - 0\cdot 12/Z. \tag{30}$$

Similarly we find

$$\int_0^{T_0} dU/T = 0\cdot 69 - 0\cdot 42/Z. \tag{31}$$

The case of $r = 1/4$, $Z = 12$, is represented in fig. 4. AGD represents the value of $(nk)^{-1}\partial\phi/\partial\sigma$ for superlattice order according to the simple treatment, and OF its value for liquid order according to (27). The intersection of the two curves takes place at a much higher value of σ than for $r = 1/2$, $Z = 6$, and the area of the loop OGA is actually nearly equal to the whole area between the two curves beyond G. If OGF accurately represented liquid order this would mean that practically complete order of neighbours ($\sigma = 1$) could be realized without long-distance order setting in at all. We can see in a general way that this must be a tendency for small r. For when the alloy has only a small fraction of atoms of one kind,[*] A atoms, say—or when Z is small—it is evidently possible for *all* the *nearest* neighbours of *all* A atoms to be B atoms without having long-distance order. In that case long-distance order will not be established at all by the interaction of nearest neighbours. The equilibrium state of the alloy at absolute zero is then one in which the A atoms, with their shells of B atoms, are randomly distributed—this having greater entropy, but no more energy than the particular state in which the A atoms form a superlattice. For such small r long-distance order is only established by the interaction of the A atoms with atoms *outside* the first shell.

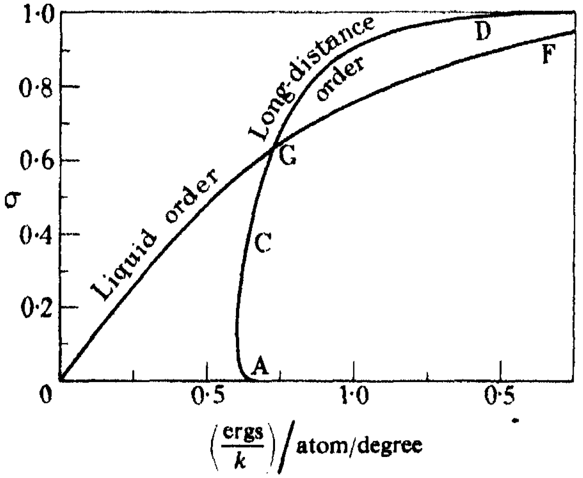


FIG. 4.

The high intersection of the two curves in fig. 4 shows that already for $r = 1/4$ there is a marked tendency towards the establishment of a high degree of order of nearest neighbours without having long-distance order. Actually the value of 1/4 for r is, however, not sufficiently small to make it possible to arrange the atoms so as to have *complete* order of neighbours without having long-distance order. There must therefore be a critical point at which transition to long-distance order takes place. This would indicate that the correct curve representing local order bends over towards the axis of abscissae somewhat more than the curve OGE, which represents local order according to the approximate formula (27).

---

[*] In this section we assume $a = r$.

Finally, it might be pointed out that the higher intersection of the curves in fig. 4 than in fig. 3 means that for the values of Z concerned the simple treatment of Bragg and Williams does not provide so good an approximation to the solution of Bethe's problem (with interaction only between nearest neighbours) for $r = 1/4$ as it does for $r = 1/2$. For $r = 1/4$, Z would have to be about 40 in order that the simple treatment should provide as good an approximation as it does with $Z = 6$ for $r = 1/2$.

In concluding this section it should be pointed out that if there is appreciable direct interaction between atoms which are not nearest neighbours, the simple treatment will diverge less from the actual state of affairs than it does from the exact solution of the problem assuming interaction only between nearest neighbours. The latter problem is one to which the simple treatment is least applicable, and it is in a way surprising that there are cases such as $r = 1/2$ for which it provides as good a solution of it as it does.

7—GENERAL DISCUSSION

The principal feature of order-disorder transformation in alloys is the existence of a critical temperature. This receives a natural explanation if, *as in the theory of Bragg and Williams, we recognize that the forces tending* to produce order increase with the degree of order.

Next to giving a critical temperature the most general theoretical result concerns the total energy of disorder given out by the alloy as it is cooled, or more precisely the value of the integral

$$\int_{T=0}^{\infty} dU(T)/T,$$

U (T) denotes the energy of disorder for the equilibrium state at temperature T. Since $dU(T)/T = d\phi$* we have

$$\int_0^{\infty} dU(T)/T = \phi(\infty) - \phi(0) = \phi_d, \text{ say,} \tag{32}$$

$\phi(\infty)$ denotes the entropy of the completely disordered atomic arrangement at $T = \infty$, and $\phi(0)$ is the entropy of the ordered arrangement at $T = 0$.

---

* $d\phi = dQ/T = (dU + pdv)/T$. In these changes pdv is negligible so that $d\phi = dU/T$.

The principal assumption upon which the value of ϕ_d depends is that the phenomena concerned *are* due to changes in atomic arrangement. There is also the auxiliary assumption that the ordinary entropy of vibration of the atoms is independent of the changes in atomic arrangement.

On these assumptions we find for the general case

$$\begin{aligned}\int_0^\infty d\text{U}(\text{T})/\text{T} = \phi_d &= -nk\{(1-r)\log(1-r) + a\log a \\ &\qquad\qquad - (a-r)\log(a-r)\}, \quad a > r \\ &= -nk\{(1-a)\log(1-a) + r\log r \\ &\qquad\qquad - (r-a)\log(r-a)\}, \quad a < r. \qquad (33)^{*\dagger}\end{aligned}$$

For $a = r = 1/2$ this gives

$$\int_0^\infty d\text{U}(\text{T})/\text{T} = nk\log_e 2 = 0{\cdot}69\,nk. \qquad (34)$$

In the B-W type of treatment the alloy is completely disordered above the critical temperature so that $\phi(\text{Tc}) = \phi(\infty)$, and therefore

$$\int_0^{\text{T}_c} d\text{U}(\text{T})/\text{T} = \int_0^\infty d\text{U}/\text{T} = 0{\cdot}69\,nk. \qquad (35)$$

In Bethe's treatment there is some degree of order even above T_c, but for $r = 1/2$ this is small, and actually for $\text{Z} = 8$ (equation 31)

$$\int_0^{\text{T}_c} d\text{U}(\text{T})/\text{T} = \phi(\text{T}_c) - \phi(0) = 0{\cdot}64\,nk. \qquad (36)$$

This differs very little from the value according to the B-W type of treatment. Since these two treatments represent, in a sense, extreme assumptions regarding the nature of the ordering forces,‡ the value of $\int_0^{\text{T}_c} d\text{U}/\text{T}$ must, on the above basis assumption, be close to $nk\log 2$, whatever kind of theory is used to work out the details of the phenomenon. Since the energy of disorder is nearly all released in a region of temperature which does not extend much below T_c,

---

* If n is equal to Loschmidt's number then nk is equal to the molecular gas-constant $\text{R} \approx 2$ calories/°.

† It will be noticed that this formula gives a "peaked" maximum at the simple composition $a = r$.

‡ The solution given by the B-W type of treatment is *exact* if the "ordering" force acting on an atom is uniquely determined by the superlattice order (which implies interaction over large distances). Bethe's treatment, on the other hand, is exact if this force is determined by the nearest neighbours of the atom concerned.

$$\int_0^{T_c} dU/kT \approx U(T_c) - U(0)/kT_c\,, = \rho. \tag{37}$$

The calculated value of ρ for $r = 1/2$ is compared with experiment in Part II, the alloy concerned being CuZn. The agreement is good. Actually the observed value of $\int_0^{T_c} d/UT$ is about 0·5. The principal basic assumption of the theory is thus substantiated. This agreement is, in fact, the simplest and most direct quantitative evidence that the phenomena dealt with are due to the formation of a superlattice from a state of disorder. It should be pointed out that the alloy CuZn, for which the comparison with experiment has been made, particularly lends itself to this test of the theory. In the first place it belongs to the case of $r = 1/2$, so that in Bethe's treatment, as well as in that of Bragg and Williams, there is no question of much ordering above the critical temperature. Secondly it has a much smaller time of relaxation at the critical temperature than most alloys, and this makes it possible to realize a high degree of order in the annealed alloy.*

For $r = 1/4$ the value of $\int_0^{T_c} dU/T$ in the theory of Bragg and Williams, is $\phi_d = nk\,(\frac{3}{4}\log\frac{4}{3} + \frac{1}{4}\log 4) = 0{\cdot}56\,nk$. The observed value according to Sykes's measurements for Cu_3Au (§ 5, Part II) is about $0{\cdot}50\,nk$. The agreement is again good. The observed value is, in fact, probably a little too low owing to the effects of the finite rate of relaxation, and the actual value may even be slightly greater than $0{\cdot}56\,nk$. In Bethe's theory $\int_0^{T_c} dU/T$ is less than ϕ_d owing to the local order still present above T_c. For $r = 1/2$ the difference is only about 10%, but for reasons given at the end of the preceding section it may be more important for $r = 1/4$. In that case, it would appear that the value of $\int_0^{T_c} dU/T$ on Bethe's theory is somewhat less than the actual value. This would indicate that local order of neighbours may not be so important for $r = 1/4$, as it is assumed to be in Bethe's type of treatment, and that the B-W treatment is more appropriate. We shall not dwell further on this point, since further investigations are in progress. On the theoretical side the detailed requirements of Bethe's theory for $r = 1/4$ are being worked out by Dr. R. Peierls in this laboratory, and Dr. Sykes is carrying out more exact measurements on the heat content.

---

* It will be seen that the observed value of $\int_0^{T_c} dU/T$, viz., 0·5, is actually somewhat less than the theoretical value which is in the region of 0·6–0·7. The difference may be due partly to the annealed CuZn in Sykes's experiments not being completely ordered at the lower temperature owing to some disorder being frozen in, and also partly to variations in the entropy of vibration, ϕ^{vibs}, with degree of order. This would mean a partial failure of the auxiliary basic assumption referred to above. That ϕ^{vibs} does not change appreciably with order is, however, indicated by the small changes in volume which accompany changes in order.

After the question of the value of $\int_0^{T_c} dU/T$ comes that of the rate at which the energy of disorder is released when the alloy is cooled below its critical temperature. This is much more dependent on detailed assumptions regarding the ordering forces, and it is not unexpected that in this respect the divergence from experiment is rather pronounced. Actually, the observed rate at which the energy of disorder is given out is shown in Part II to be considerably greater than that calculated, whether we take the results obtained by Bethe, using the order of neighbours σ as the parameter which determines the energy, or the results given by the B-W type of treatment based on the superlattice order S as parameter. In fact, the difference between the results of the two treatments is much less than their common divergence from experiment. This indicates that the discrepancy cannot be removed by some intermediate kind of treatment, such as, for instance, a treatment in which Bethe's method is extended to take into account the direct interaction between next-nearest neighbours as well as that between nearest neighbours. It would appear, therefore, that the necessary modification of the theory, whatever it is, is one which can be considered on the basis of either treatment.

In the B-W treatment the rate of release of energy is determined by the rate at which the intersection of PQ, fig. 3 (representing $T^{-1}\partial U/\partial\sigma = T^{-1}\partial U/\partial s^2$ in the B-W theory), with AGD (representing $\partial\phi/\partial s^2$) runs up along AGD as PQ moves to the right with decreasing temperature.* In Bethe's treatment it is the intersection of PQ with OCD that is concerned. It is evident that in both cases these rates would be considerably increased if PQ sloped a little to the right instead of being vertical, *i.e.*, if V_0, in the B-W relation $V = V_0S$, or V_B in Bethe's treatment, increased with the degree of order. As a measure of the rate of release of the energy of disorder, we shall take two quantities, the ratio π of the abnormal specific heat just below T_C to the normal specific heat and the ratio λ of the energy released between T_C and $0 \cdot 9$ T_C to the total energy released. Then, assuming, as hitherto, that V_0 and V_B are independent of the atomic arrangement, we have

---

* To represent the treatment of Bragg and Williams in terms of the notation of that treatment σ for these curves is to be translated into S^2, and V_B to $V_0/2Z$.

$$\pi_{B-W} = 0\cdot50 \qquad \lambda_{B-W} = 0\cdot28$$

$$\pi_{Bethe} = 0\cdot60 \qquad \lambda_{Bethe} = 0\cdot39.$$

According to Sykes's observations on CuZn (*see* § 4 of the Part II paper)

$$\pi_{obs} = 2\cdot5 \qquad \lambda_{obs} = 0\cdot6.$$

Now, if we assume that $V = V_0S = V_0^0S\,(1 + gS^2)$ and $V_B = V_B^0\,(1 + g\sigma)$ then for $g = 0\cdot24$, for instance,*

$$\pi'_{B-W} = 1\cdot8 \qquad \lambda_{B-W} = 0\cdot48$$

$$\pi'_{Bethe} = 4^{\dagger} \qquad \lambda'_{Bethe} = 0\cdot57.^{\dagger}$$

We thus see that, by assuming a small increase of V_0 or V_B with the degree of order, it is possible to bring the calculated values into general agreement with experiment. Though this assumption is *ad hoc*, it appears to be the most feasible one to make. A more rapid rate of ordering below the critical temperature could, of course, be achieved by making V_0 or V_B increase with decreasing temperature. A variation with T^{-1}, for instance, would double the rate of release of energy of disorder. There is, however, no reason to expect such a dependence on temperature. It might also be suggested that a modification of the curves representing $\partial\phi/\partial s^2$ or $\partial\phi/\partial\sigma$ be made instead of a modification of the curves representing $\partial U/\partial s^2$ and $\partial U/\partial\sigma$ (*i.e.*, V_0 and V_B). In the first place, however, neither of the former is likely to involve any appreciable errors of computation. In fact, $\partial\phi/\partial s^2$ is calculated without approximation, and though the calculation of $\partial\phi/\partial\sigma$ by Bethe is not exact, the closeness of the results of his first and those of his second approximations indicates that any errors of computation are small. The other possible source of error in the expressions used for $\partial\phi/\partial s^2$ and $\partial\phi/\partial\sigma$ is that the ordinary entropy of vibration of the atoms may depend on the degree of order. Though this dependence is unlikely to be appreciable, its effect on the rate of ordering may be more marked than on the total energy of disorder. Even so, it appears at the moment

---

[*] A transition with latent heat would be realized if g was a little greater. In the B-W treatment, for instance, the critical value of g for latent heat is 1/3.

[†] These values have been calculated assuming $\partial\phi/\partial\sigma$ to be the same as in fig. 3. This is not quite correct, owing to local fluctuations in σ, and therefore in V_B. It would appear that a rigorous treatment would, for a given value of g, give still greater values of π and λ, and thus only strengthen the point which it is desired to make: that only a small dependence on σ is necessary to considerably increase π and λ.

that a small dependence of V_0 or V_B on the degree of order is the most likely cause of the existing discrepancies.

In a theoretical treatment of the dependence of the ordering " energies " on the degree of order, it will, of course, be necessary first to settle the question of whether the order of neighbours σ is always a more relevant parameter than the superlattice order S; in other words, whether the forces tending to keep an atom in its place are determined by the nature of its nearest neighbours or determined more directly by the superlattice order. Bethe considers only the first of these alternatives. It is, however, not inconceivable that in certain cases the second may be the more correct. It is known that the energy of the free electrons in a metal or an alloy is closely connected with the distribution of the reflecting planes formed by the atoms, and in certain cases the introduction of new reflecting planes—which takes place when superlattice order sets in—may have a larger effect on the energy of the alloy than the direct effect of the corresponding increase in the order of neighbours. These are questions which await solution.

The writer wishes to take this opportunity of thanking Professor W. L. Bragg for his very encouraging interest and help in the work, and of thanking Dr. R. Peierls for many helpful discussions on the subject.

SUMMARY

The equilibrium equation for a general case, assuming the energy associated with the atomic arrangement to be uniquely determined by the superlattice order, is derived from the Boltzmann distribution formula and also from the principle of minimum free energy. The entropy and energy corresponding to the solution recently given by Bethe, assuming the energy of the atomic arrangement to be determined by the order of nearest neighbours, is also discussed, and a method of estimating the quantitative relation of Bethe's type of treatment to that of Bragg and Williams is given. The paper concludes with a general discussion of existing theory and of the possible modifications for improving the relation to experiment.

The Pressure of a Degenerate Electron Gas and Related Problems

By Sir ARTHUR EDDINGTON, F.R.S.

(*Received June* 14, 1935)

1—The " ordinary " formula for the minimum electron pressure P corresponding to an electron density σ is of the form $P = K\sigma^{\frac{5}{3}}$. It has been generally accepted that this is a non-relativistic approximation, applying only to slow-moving electrons; and a " relativistic " formula has been given, intended to take account of change of mass with velocity. In a recent paper* I have contended that the " ordinary " formula is the exact relativistic solution of the problem, and that the " relativistic " formula rests on a misconception. Since a decision on this point has far-reaching consequences in the theory of dense stars, and, moreover, has a fundamental bearing on the union of quantum theory with relativity theory, it has seemed desirable to supplement my earlier paper.

I think the earlier paper makes it clear that whether the " relativistic " formula is right or wrong, existing proofs of it cannot be accepted. Briefly, the fallacy lies in the fact that the cell-division of phase space is obtained by *standing waves*, but these are assigned an energy which has been derived for *progressive waves*. If this is not simply a mistake, it is a step which requires careful justification; and up to the present no one seems to have given any reason why one should assume that the energy or hamiltonian of standing waves is the same as that of progressive waves. We may recall that the energy of plane progressive waves is easily calculated, because by a Lorentz transformation they are converted into waves advancing along the time-axis, and the energy is then the rest-energy of the particle; if the Lorentz transformation is applied to standing waves no such simplification results, and the reduction of the particle to rest involves an altogether different type of transformation.

In § 2 I give what seems to be a straightforward deduction of the formula $P = K\sigma^{\frac{5}{3}}$ from relativistic principles. In this proof the pressure is found directly, and not *via* the hamiltonian.

It is next necessary to meet the criticism that the result is widely subversive of accepted formulae throughout the whole range of quantum theory, and so indirectly comes into conflict with observation. It is shown in §§ 3–6 that no such conflict arises, and that the usual (Dirac) theory of atomic phenomena follows logically from the present basis.

---

* 'Mon. Not. R. Astr. Soc.,' vol. 95, p. 194 (1935).

These results help to make clear the connexion between relativity theory and wave mechanics. In §§ 7–13 they are applied to a determination of the value of the cosmical constant—a problem which I had hitherto only been able to treat approximately. Light is thrown on many side problems. The existence of a vast number of occupied negative energy levels is one of the essential consequences of the analysis.

2—In macroscopic theory, energy, momentum, and pressure are components of an energy tensor $T_{\alpha\beta}$. It is part of the definition of these quantities that they satisfy the law of conservation $(T^{\alpha\beta})_{\beta} = 0$; and by this property $T_{\alpha\beta}$ is identified with a geometrical tensor

$$-8\pi\kappa T_{\alpha\beta} = G_{\alpha\beta} - \tfrac{1}{2}g_{\alpha\beta}(G - 2\lambda), \tag{2.1}$$

which satisfies the law of conservation identically.

Here λ may be any constant, a change of λ being merely a change of the zero from which energy and pressure are reckoned. But the zero condition ($T_{\alpha\beta} = 0$) should not be taken to correspond to entire absence of matter and radiation, or in wave mechanics to zero probability of the presence of particles and photons. For if we adopt this as zero condition everywhere, we are comparing the actual universe with a universe completely devoid of matter and radiation and therefore without observable characteristics. We cannot use as a reference standard for physical measurements an ideal system in which all such measurements become indeterminate.

In equations intended to apply throughout all space, it is preferable to choose λ so as to correspond to the actual amount of matter in the universe; so that the passage from the actual to the zero condition involves no creation of matter but only a redistribution. In ideal problems with artificial boundary conditions there is no criterion for fixing λ. In such problems the energy tensor contains an arbitrary additive term $-\lambda g_{\alpha\beta}/8\pi\kappa$.

To make the transition to wave mechanics, $T_{\alpha\beta}$ must be interpreted as the expectation value (for unit volume) of an operator $\mathbf{T}_{\alpha\beta}$. The only appropriate operation is differentiation; and to satisfy tensor conditions we must set

$$\mathbf{T}_{\alpha\beta} = -\frac{1}{m}\frac{\delta^2}{\delta x_{\alpha}\,\delta x_{\beta}} \tag{2.2}$$

where δ stands for covariant differentiation and m is an invariant. We shall be concerned only with scalar wave functions, so that (in natural co-ordinates) the covariant derivatives are ordinary derivatives. The pressure P is $\frac{1}{3}(T_{11} + T_{22} + T_{33})$, so that

$$P = \text{expectation value of } -\nabla^2/3m. \tag{2.25}$$

The connexion between macroscopic theory and wave mechanics has commonly been made through the momentum vector*

$$\mathbf{p}_\alpha = -i\delta/\delta x_\alpha \qquad p^\alpha = m\,dx_\alpha/ds. \tag{2.3}$$

Substituting in (2.2) this gives $\mathbf{T}_{\alpha\beta} = \mathbf{p}_\alpha \mathbf{p}_\beta/m$, so that

$$T^{\alpha\beta} = m\,\frac{dx_\alpha}{ds}\frac{dx_\beta}{ds},$$

which is a well-known form for a system of particles when the right-hand side is summed in an appropriate (invariant) way. Thus there is no conflict between (2.2) and (2.3); but the former expresses the connexion more directly and more generally.

The tensor $T_{\alpha\beta}$ refers to unit volume. The expectation value of $\mathbf{T}_{\alpha\beta}$ for each particle is its eigenvalue for the particle multiplied by the probability that the particle is in the unit volume. In general it is a subtle problem to define what fraction of a moving particle is to be included in a given unit volume. But in the application referred to in § 1 this difficulty does not arise. We deal with a rectangular block of matter (in the interior of a star) of dimensions $l_1 \times l_2 \times l_3$, containing $\sigma l_1 l_2 l_3$ electrons. We take axes such that the block as a whole is at rest, and we analyse the internal state of the block into standing waves in the relative co-ordinates ξ_1, ξ_2, ξ_3. By the exclusion principle there are not more than two electrons to each wave. Our purpose is to calculate the minimum value of P; so that by (2.25) we must find the $\frac{1}{2}\sigma l_1 l_2 l_3$ waves which give the lowest eigenvalues of $-\nabla^2$. We shall then have

$$l_1 l_2 l_3 T_{11} = \Sigma \mathbf{T}_{11}, \text{ etc.,} \tag{2.4}$$

where the summation applies to the $\sigma l_1 l_2 l_3$ electrons which by their wave specification are wholly and permanently within the block.

The scalar wave function ψ is analysed (within the block) into Fourier components, viz., $\psi = \Sigma\,\psi_{n_1 n_2 n_3}$ where

$$\left.\begin{aligned} \psi_{n_1 n_2 n_3} &= a_{n_1, n_2, n_3} \cos\frac{2\pi n_1 \xi_1}{l_1} \cos\frac{2\pi n_2 \xi_2}{l_2} \cos\frac{2\pi n_3 \xi_3}{l_3} \\ \psi_{n_1, n_2, -n_3} &= a_{n_1, n_2, -n_3} \cos\frac{2\pi n_1 \xi_1}{l_1} \cos\frac{2\pi n_2 \xi_2}{l_2} \sin\frac{2\pi n_3 \xi_3}{l_3} \end{aligned}\right\} \tag{2.5}$$

---

* We employ units such that the constants c, $h/2\pi$ are unity. The constant of gravitation κ is therefore not unity, and is retained in (2.1).

and n_1, n_2, n_3 are integers (positive in (2.5)). The factors can be cosines or sines, and (as a convention) we distinguish sine factors by negative values of n. The condition governing this analysis is that the component wave functions must constitute a complete set of orthogonal functions for a distribution of ψ inside the block.

Let

$$\varpi_\alpha = 2\pi n_\alpha / l_\alpha. \tag{2.6}$$

Taking ϖ_1, ϖ_2, ϖ_3 as rectangular co-ordinates, each component wave is represented by a point in ϖ-space. The points form a rectangular lattice, and their density in ϖ-space is $l_1 l_2 l_3/(2\pi)^3$. Hence the $\frac{1}{2}\sigma l_1 l_2 l_3$ waves will occupy a minimum volume $\frac{1}{2}(2\pi)^3\sigma$. If this volume is a sphere of radius r, we have

$$\tfrac{4}{3}\pi r^3 = \tfrac{1}{2}\sigma(2\pi)^3. \tag{2.7}$$

Writing $\varpi^2 = \varpi_1{}^2 + \varpi_2{}^2 + \varpi_3{}^2$, the mean value of ϖ^2 for the sphere is

$$\overline{\varpi^2} = \tfrac{3}{5} r^2 = \tfrac{3}{5}\left(\frac{3}{8\pi}\sigma\right)^{\frac{2}{3}} (2\pi)^2 \tag{2.8}$$

by (2.7)

By (2.5), $\nabla^2 = -\varpi^2$. Hence, summing for the σ particles in unit volume,

$$P = \Sigma(-\nabla^2/3m) = \tfrac{1}{3}\Sigma\varpi^2/m = \tfrac{1}{3}\sigma\overline{\varpi^2}/m \tag{2.85}$$

$$= \tfrac{1}{5}\left(\frac{3}{8\pi}\sigma\right)^{\frac{2}{3}} \frac{(2\pi)^2}{m}\sigma^{\frac{5}{3}}. \tag{2.9}$$

To convert to C.G.S. units the factor $(2\pi)^2$ is replaced by h^2, our theoretical units having been chosen so that $h/2\pi = 1$. Equation (2.9) is then the "ordinary" degeneracy formula which has usually been condemned as non-relativistic. But it is here obtained directly from the relativistic equation (2.2).

3—We shall now deduce the hamiltonian for these waves, that is to say the expression for T_{44} in terms of T_{11}, etc. This is not required in the calculation of P, but it is of importance for further developments. We write δT_{44} for the part of T_{44} due to the standing waves which furnish P, and set

$$P = \Sigma P_r \qquad \delta T_{44} = \Sigma (T_{44})_r, \tag{3.1}$$

where P_r and $(T_{44})_r$ are the contributions of the rth wave. The same P can be constituted out of individual waves in many ways. We first notice that two sets of waves which give the same δT_{44} must also give the same P; for, if not, the change from one to the other would create an extra pressure ΔP with no change of the other components of $T_{\alpha\beta}$. It is known from the theory of gravitational waves that a gravitational source ΔP, unaccompanied by a source in some other component, is impossible.*

Hence, if $(T_{44})_r = f(P_r)$, the function f must be such that $f(a) + f(b) = f(a') + f(b')$ when $a + b = a' + b'$. This condition gives $f(P_r) = \alpha P_r$, where α is a constant. Thus

$$(T_{44})_r = \alpha P_r = \tfrac{1}{3}\alpha\,(\varpi^2/m)_r. \qquad (3.2)$$

This corresponds to a hamiltonian of Schrödinger type

$$\text{energy} = \text{constant} + \tfrac{1}{2}\varpi^2/m \qquad (3.3)$$

and not to the so-called relativistic hamiltonian which is a non-linear function of ϖ^2. Comparing (3.2) and (3.3), it would seem that $\alpha = 3/2$; but we shall see in § 6 that this not a correct identification.

In one sense (3.2) is an approximation, inasmuch as, the gravitational equations being non-linear, T_{44} cannot strictly be analysed into additive contributions of the individual waves. The non-linear terms would be regarded as mutual gravitational energy of the particles represented by the waves. Whether (3.2) is an approximation or not depends on whether $(T_{44})_r$ is intended to include this potential energy. But in any case this is not the point at issue. The coefficient α may or may not be a function of the curvature of space-time and therefore of the total pressure P. Our result is that, when P is fixed, α is independent of P_r, whereas in the same conditions the "relativistic" hamiltonian makes $(T_{44})_r$ a non-linear function of P_r.

The result of this section agrees with the current conclusion, viz., that the "ordinary" degeneracy formula implies a Schrödinger hamiltonian for the waves—so that the two stand or fall together.

The method of proof used in my earlier paper was to obtain the hamiltonian for a relative (internal) wave function by considering the transformation from absolute to relative co-ordinates. Since this turned out to be of the Schrödinger type, the "ordinary" degeneracy formula followed.

4—The result of the foregoing discussion is to reinstate the Schrödinger form of hamiltonian as correct for the problem of standing waves in a block of material. It is not suggested that the Schrödinger hamiltonian is to be reinstated in all problems, *e.g.*, in the interior of an atom. If it is asked why the same considerations do not apply to the interior of an atom, we answer that they do not even apply to a distribution of matter in a smooth spherical vessel. Besides the modification of the orthogonal functions ψ_n to correspond to a spherical boundary a new feature appears. There are steady states in which the distribution has an angular momentum relative to the sphere.

---

* 'Proc. Roy. Soc.,' A, vol. 102, p. 268 (1922). We refer to the condition $h_{22} + h_{33} = 0$ corresponding to the condition $\Delta(T_{22} + T_{33}) = 0$, so that $\Delta P = 0$.

With a spherical boundary, unidirectional motion is no longer inconsistent with a steady state. Rotation in one direction corresponds to progressive waves in the angular co-ordinate θ. These are of the same type as the progressive waves associated with a particle moving freely in space, *i.e.*, rotating about the centre of curvature of space.

I have pointed out the error of applying the hamiltonian of progressive waves to standing waves. It would be an equal error to apply the Schrödinger hamiltonian of standing waves to unidirectional motion. In all problems in which there exist integrals of angular momentum Schrödinger's hamiltonian ceases to apply; and my investigation raises no suspicion that the accepted hamiltonians for these problems are incorrect.

The advent of progressive waves into a problem depends on the existence of relativity transformations of the co-ordinates, which occasion a degeneracy of the steady state solutions. In the problem of the rectangular block there are no relativity transformations of the internal co-ordinates ξ_μ; no other set of axes is " equivalent " to that which is parallel to the edges of the block. In the spherical problem all orientations of the internal rectangular axes are equivalent; so that there are three relativity transformations, viz., rotations in the three co-ordinate planes. The time-direction is still unique, being defined by the motion of the spherical boundary, so that there are no Lorentz transformations. For this limited amount of relativity it is unnecessary to resort to complete Dirac wave vectors. We can take ψ to be a wave vector with two components (2-vector) and the co-ordinate planes to correspond to three Pauli matrices $\zeta_1, \zeta_2, \zeta_3$ $(\zeta_\alpha^2 = -1)$.

With 2-vectors it is not possible to express a simple numerical density σ as a product $\psi\tilde{\psi}^*$. The density has to be treated as the sum of two " pure " components

$$\tfrac{1}{2}\sigma(1 + i\zeta_\mu), \quad \tfrac{1}{2}\sigma(1 - i\zeta_\mu). \tag{4.1}$$

Each of these is idempotent (for $\sigma = 1$) and therefore may be factorized.* The non-numerical part $\tfrac{1}{2}i\sigma\zeta_\mu$ is described as density of angular momentum in the plane corresponding to ζ_μ. When the boundary conditions are such that the angular momentum in a steady state is always zero, the two components (4.1) are constrained to have equal probability, and there is no need to separate them. But when, as in the spherical problem, the angular momentum may have any value, they have independent probabilities and constitute, therefore, distinct elementary states.

---

* 'J. Lond. Math. Soc.,' vol. 8, p. 145 (1933).

For a particle moving freely in space the relativity transformations are more general, and Lorentz transformations are included. This additional degeneracy requires a second set of Pauli matrices θ_μ, commuting with the ζ_μ. The two components of opposite spin $\frac{1}{2}\sigma\ (1 \pm i\zeta_\mu)$ are in turn analysed into components of opposite charge $\frac{1}{4}\sigma\ (1 \pm i\zeta_\mu)\ (1 \pm i\theta_\mu)$. The wave vectors now have four components, and the double set of Pauli matrices is more conveniently replaced by a set of Dirac matrices with the same non-commutative properties. In my notation the substitution is $\zeta_\mu = \mathrm{E}_{23}, \mathrm{E}_{31}, \mathrm{E}_{12}$; $\theta_\mu = \mathrm{E}_{45}, -\mathrm{E}_5, \mathrm{E}_4$.

The dissection $\frac{1}{4}\sigma\ (1 \pm i\zeta_\mu)\ (1 \pm i\theta_\mu)$ provides four sub-waves (two electrons and two protons) for each of the elementary waves (2.5). Although the dissection is not required in the problem of the rectangular block—the four sub-waves being constrained to have the same amplitude—it must be imagined to exist in order that the results may be expressed in the same form as those of problems in which unidirectional motion is not excluded.

This brief outline of the way in which the ordinary relativity theory of free electrons and of angular momentum is reached from the present starting-point will, I think, show that we have introduced no contradiction with atomic physics. It is necessary to emphasize that " relativity " is not a formula but a theory; and that its transformations are not to be applied mechanically but when relativity theory indicates that they are applicable. In the problem of the rectangular block the hamiltonian of the internal motion is not invariant for Lorentz transformations—a result foreseen by relativity theory as well as by common sense.

The fictitious particle, which temporarily replaces a group of two protons and two electrons when scalar wave functions are employed, will be called a *neutral particle.* The term has no connexion with the neutron.

5—From one aspect the rectangular volume is occupied by the set of standing waves treated in § 2. From another aspect it is occupied by n particles moving in various directions; in wave mechanics these are represented by progressive waves which combine into an n-tuple wave function. I think it will save some confusion of thought if we notice that, whereas the standing wave function represents a system in equilibrium, the n-tuple wave function represents a rapidly dispersing system; if we substitute in it a later value of the time, it gives the distribution which would be reached if each particle continued to move with its present momentum.

Every time a particle passes the boundary, the n-tuple wave function changes discontinuously. It is only by forming a time-average that we convert it into a function representing steady conditions. But in forming the time-average we must first replace the discontinuities by a continuous change; otherwise the function is not differentiable, and is useless for computing energy. The energy $(- i\partial/\partial t)$ of the smooth wave function will be altogether different from that of the instantaneous n-tuple wave function.

The general effect of this time-averaging of the discontinuities can be seen by considering the combined wave function of the rectangular volume A and the surrounding matter B which is holding it in equilibrium. The steady state (represented by the standing waves) results from the fact that, on the average, for every particle passing from A to B a similar particle passes from B to A. In the combined wave function of A and B this is merely a nominal interchange. The co-ordinates x_p of the pth particle of system A have become inappropriate to system A, and the co-ordinates $x_{q'}$ of the qth particle of system B have become inappropriate to system B; we therefore re-label the particle at x_p as the qth particle of system B, and the particle at $x_{q'}$ as the pth particle of system A. By the Fermi-Dirac rule this interchange of labels reverses the sign of the wave function. In time-averaging the succession of discontinuous reversals of sign is replaced by a continuous factor $e^{ip_i t}$ which reverses the sign at intervals $t = \pi/p_i$. When we form the energy $- i\partial/\partial t$, this factor gives an additional energy p_i. This interchange energy does not appear in the instantaneous n-tuple wave function; but it belongs inalienably to the steady system which the standing wave functions represent.

Interchange energy was, I think, first introduced in my investigation of the charge of an electron,* and I follow the interpretation of it which I then gave. It is potential energy. The system occupying the rectangular volume may be kept in equilibrium either by interchange of particles at the boundary or by a force at the boundary (exerted by the matter beyond) turning back the particles as they reach it. But, since we never know which particle is which, it is meaningless to enquire whether an old particle is turned back at the boundary or a new particle comes in. Thus the interchange effect and the force are identical; and so also are the interchange energy and the potential energy due to the field of force.

The problem of waves in a rectangular volume is often treated by assuming a wall of infinite potential at the boundary.* But the problem is more manageable if the field of force restraining the particles from dispersal is continuous throughout the volume. This represents much more closely the actual conditions (to which the results are generally applied), in which there is nothing distinctive about the boundary of the volume considered, and the collisions or interchanges which prevent dispersal of the system occur evenly through the volume. The corresponding interchange energy or potential energy is then evenly distributed.

---

* 'Proc. Roy. Soc.,' A, vol. 122, p. 358 (1928).

6—Since the energy of the standing waves includes a certain amount of potential energy it would be wrong to equate it to a hamiltonian which represents only the kinetic energy of a dispersing system. When the velocities of the particles are small, the Schrödinger and " relativistic " hamiltonians agree. But these represent only kinetic energy; so that even in the limit, when the classical approximations suffice, they will not be equal to our hamiltonian which includes the potential energy inalienably associated with standing waves. We shall show, however, that in the classical approximation the kinetic energy and potential energy are equal; thus in the limit our hamiltonian is twice as great as the others. This fixes the value of α as 3, instead of 3/2 as first suggested in §3.

There is nothing recondite about the waves in (2.5). They are simply sound waves. The wave functions ψ are real, so that the conjugate wave functions $\overline{\psi}$ are the same. The particle density is therefore $\sigma = \overline{\psi}\psi = \psi^2$. All we have done is to perform a Fourier analysis of $\sqrt{\sigma}$, and, until we introduce the exclusion principle, the whole investigation comes under the ordinary theory of sound. In the classical approximation the energy of standing waves of sound is half kinetic and half potential; and we have used this result to fix $\alpha = 3$. Hence

$$\delta T_{44} = 3P \qquad \delta T = 0, \tag{6.1}$$

since in rectangular co-ordinates $T = T_{44} - T_{11} - T_{22} - T_{33}$.

It is well known that $T = 0$ for electromagnetic waves. Accordingly when radiation is absorbed and converted into increased energy of molecular motion of the kind here treated, there is no change of T. This is subject to the limitations (1) that the volume of the gas is kept unchanged, and (2) that no unidirectional motion, *e.g.*, atomic or molecular spin, is produced.

---

* According to R. H. Fowler, "Statistical Mechanics," p. 535, an infinite wall of potential gives the boundary condition $\psi = 0$. He assumes that the potential is a simple scalar; but as the field becomes large this Newtonian type of approximation must break down, and it is difficult to see what the investigation proves. Whether correct or not, the result can refer only to gas contained in a vessel of dimensions small compared with the free path of the molecules.

The condition $\delta T = 0$ also agrees with elementary dynamics in associating no rest-mass with the relative (internal) motion of the parts of a system. When the motion of two particles of masses m_1, m_2 is resolved into an external motion of a particle of mass $m_1 + m_2$ occupying the centre of gravity and an internal motion of a particle of mass $m_1m_2/(m_1 + m_2)$ in the relative orbit, the whole rest-mass is associated with the external motion. Similarly in wave mechanics the rest-mass belongs to the energy tensor of the external wave function, and not to the wave function of the internal motion studied in § 2 which contains the pressure.*

Suppose that we have a vessel containing monatomic gas and raise the temperature of the gas by adding heat energy H, what will be the change of the mass of the vessel, measured in the ordinary way by its acceleration under a given applied force? The ordinary answer would be H/c^2; but our formula (6.1) suggests that it should be $2H/c^2$. It might be suggested in defence of $2H/c^2$ that we have to take into account the gravitational and inertial effect of the increased tension in the walls of the vessel; but I think there is no doubt that H/c^2 is at least approximately correct. The reason why (6.1) fails in this case is that the tension in the walls of the vessel represents interatomic processes—electrical or spin effects—which, as seen in § 4, introduce unidirectional waves.

I emphasize these elementary illustrations, because the fixed relation between changes of energy and changes of pressure must ultimately rest on convention, and it is not a question of discovering a natural law but of ascertaining current practice and definition. The energy tensor contains an additive term with arbitrary coefficient λ (§ 2), and the condition $\delta T = 0$ is equivalent to defining the choice of λ. Our result that $\delta T = 0$ when a quantity of radiation is absorbed and its energy converted into standing material waves may be regarded as a principle of stationary action for small variations of the equilibrium state of radiation and matter. Here, as usual, the action principle is of the nature of a definition of the quantities concerned in it.

7—Curvature of space-time is equivalent to a distribution of ψ-waves, since both devices are employed for the same purpose—to represent the energy and momentum of physical systems. It is not supposed that either curvature or ψ-waves are *objective*. I have obtained an approximate relation between the two representations in earlier papers; I then had in mind progressive waves. For standing waves the problem of finding the exact connexion seems to be very much easier, and I think the following is a complete solution. The main aim is to find the relation between the natural constants used respectively in relativity theory and atomic physics; for this purpose it is sufficient if we can find *one* problem which is tractable from both points of view.

---

* 'Mon. Not. R. Astr. Soc.,' vol. 95, p. 198 (1935).

Consider a distribution of matter of uniform density in equilibrium. It will accordingly occupy part or all of a static spherical space. To introduce boundary walls limiting it to part of space would be going out of our way to create difficulties; we therefore suppose the same uniform distribution to extend throughout space. In stereographic co-ordinates the line element is*

$$ds^2 = -(1 + r^2/4R^2)^{-2}(dx^2 + dy^2 + dz^2) + c^2dt^2, \tag{7.1}$$

and the pressure and density are found to be

$$\left.\begin{aligned}(8\pi\kappa/c^4)\,P &= -R^{-2} + \lambda\\ (8\pi\kappa/c^2)\,\rho &= 3R^{-2} - \lambda\end{aligned}\right\}. \tag{7.2}$$

Since we are now aiming at numerical results, we shall insert the constants c, $h/2\pi$, suppressed in §§ 2–6.

Instead of treating R as a radius of curvature, we may treat

$$\beta = (1 + r^2/4R^2)$$

as a variable gauge-factor for measurements made in a flat space with Euclidean co-ordinates (x, y, z). Then by (7.1) the co-ordinate length is β times the natural length ds, and the co-ordinate density ρ_0 is β^{-3} times the uniform natural density ρ. Introducing real scalar wave functions $\psi = \bar{\psi}$, so that $\rho_0 = \psi\bar{\psi} = \psi^2$, we have

$$\begin{aligned}\psi &= \rho^{\frac{1}{2}}\,\beta^{-\frac{3}{2}} = \rho^{\frac{1}{2}}(1 + r^2/4R^2)^{-\frac{3}{2}}\\ &= \rho^{\frac{1}{2}}\left(1 - \frac{3}{8R^2}(x^2 + y^2 + z^2) + \ldots\right)\end{aligned} \tag{7.3}$$

First suppose that the density is due to the probability distribution of a single neutral particle of mass m_1, whose wave function accordingly is (7.3). The eigenvalue of ∇^2 at the origin is $-9/4R^2$. Hence by (2.25), with the natural constants inserted,

$$P = -\frac{1}{V}\left(\frac{h}{2\pi}\right)^2\frac{\nabla^2}{3m_1} = \frac{1}{V}\left(\frac{h}{2\pi}\right)^2\frac{3}{4R^2m_1}, \tag{7.41}$$

---

* *See*, for example, Tolman, "Relativity, Thermodynamics, and Cosmology," equations (138.4), (139.3), (139.4).

the expectation value (for unit volume) being formed by multiplying the eigenvalue by the probability $1/V$ that the particle is in the unit volume considered. $V =$ volume of spherical space $= 2\pi^2 R^3$. We write

$$(2\pi/h)^2 m_1 = \mu_1, \tag{7.42}$$

so that

$$P = \tfrac{3}{4}R^2\mu_1 V. \tag{7.43}$$

By (6.1) the energy density ρc^2 of a standing wave is 3P. Hence the total mass belonging to the wave is

$$m' = \rho V = 3PV/c^2 = 9/4R^2c^2\mu_1. \tag{7.5}$$

Since the standing waves imply that the particle is in motion, we cannot identify m' with the rest-mass m_1; but their relation is found as follows. Setting $\rho c^2 = 3P$ in (7.2), we find

$$\lambda = 3/2R^2 \qquad 8\pi\kappa\rho/c^2 = 3/2R^2. \tag{7.61}$$

If, however, we treat R as a curvature, so that there are no waves and consequently $P = 0$, we find

$$\lambda = R^{-2} \qquad 8\pi\kappa\rho/c^2 = 2R^{-2}. \tag{7.62}$$

The change from (7.61) to (7.62) is a change of the arbitrary constant λ in the energy tensor, *i.e.*, a change of the zero from which energy and pressure are reckoned. In elementary theory the rest-energy is defined to be the value of the energy when the other components of the energy-momentum vector vanish. This is included in the more general definition that the rest-energy is the value of the energy when the other components of the energy tensor vanish. For progressive waves, the other components can be made to vanish by applying a rotation (Lorentz transformation); for standing waves, they are made to vanish as above by altering the arbitrary constant until $P = 0$. Thus ρ in (7.62) corresponds to m_1, whereas ρ in (7.61) corresponds to m'. Hence

$$m' = \tfrac{3}{4}m_1. \tag{7.7}$$

Then by (7.5)

$$m_1 = 3/R^2c^2\mu_1. \tag{7.81}$$

And by (7.42)

$$m_1 = h\sqrt{3}/2\pi cR. \tag{7.82}$$

8—Next suppose that the density is due to a probability distribution of N′ neutral particles each of mass m_1. By the exclusion principle the particles are represented by orthogonal wave functions, each particle being assigned a different wave function. To obtain the "ground state" of the distribution we must choose the waves of lowest energy.

The wave functions or states can be classified similarly to those of an atom, except that we are not troubled by the duplication of states due to opposite directions of atomic spin.* In the atom the waves are concentrated in small volume by a controlling Coulomb field. Here the gauge-factor β fulfils the same role; or, what comes to the same thing, the waves spread to the natural limit imposed by the finitude of space. The lowest state, which we shall call the K state, represents a uniform distribution over all space; it is, of course, the state occupied by the solitary particle discussed in § 7. The remaining states are unsymmetrical in at least one dimension. We call the highest of the occupied N′ states the *limit state*, and denote its energy by m_2c^2. The limit state corresponds to very high quantum number—actually about 10^{26} —so that the energy levels have become practically continuous. For other than cosmological problems, there is a practically inexhaustible supply of particles with energies nearly equal to the limit energy.

The above analogy with the atom is obtained by considering the distribution in stereographic projection. Presumably a more symmetrical treatment would be obtained by using spherical representation; the wave functions would then correspond to spherical harmonics in four dimensions. But it is unnecessary to develop the detailed theory of these wave functions. For our purpose we require only two results: (1) the K state is a uniform distribution such as we have treated in § 7; (2) the mean energy $\overline{m}c^2$ of all the particles is $\frac{3}{5}$ of the limit energy, so that

$$\overline{m} = \tfrac{3}{5}m_2. \tag{8.1}$$

This last result was obtained in (2.8). Although it was there found for rectangular wave functions, it applies generally; since it depends only on the well-known theorem that the density of the discrete wave functions is uniform in phase space, or, as it is often expressed, they occupy equal unit cells.

9—Let us now add a particle to the system of N′ particles in the ground state. The lowest vacant state corresponds to energy m_2c^2. We must therefore endow the particle with a threshold energy m_2c^2, in addition to its visible energy available for transfer to other particles by collision, radiation, etc.

In elementary quantum theory the system under discussion is always treated as an independent addition to " the rest of the universe," which constitutes a fixed background (both material and geometrical). No notice is taken of the resulting change of curvature of space-time; no notice is taken of the N′ other particles existing in the universe, which compete for the states of lowest energy. In order to justify this independence it must be assumed that the rest of the universe is in the ground state, with no vacant levels for the added particles to drop into, and no excited particles to drop into vacant levels in the added system. In other words, quantum theory represents its systems on an *impermeable* background. This does not require that the actual universe should be in its ground state; it means that any deviation from the ground state must be mentioned explicitly and described as if it were an addition to the impermeable background supplied by the ground state.

---

* That is reserved until the N′ neutral particles are converted into 4N′ charged particles in § 11.

For example, if the added system is a vessel containing particles at temperature T, statistical equilibrium requires that the background should be excited to the same temperature. But the excited background is not treated as such; it is described as an impermeable background plus a field of radiant energy; and exchanges of energy between the added particles and the excited background are described as exchanges of energy between the added particles and the added radiation. A single vacant level in the background is described as a photon. Near the threshold energy the energy levels are almost continuous, so that photons may have practically any energy* up to the limit m_2c^2.

In non-equilibrium conditions the added particles may be localized as wave-packets. Similarly the vacant energy levels can form wave-pockets. A wave-pocket might be described as a minus-particle, but so long as we confine attention to neutral particles the wave-pocket is nothing more than the localization of a photon such as occurs, for example, in the observation of individual X-ray effects in an expansion chamber. Different considerations arise when vector wave functions are introduced and the states are divided into four sub-states representing charged particles. The charged minus-particle constituting a vacant sub-state is evidently to be identified with a positron or negatron. The rest-masses of these minus-particles must be determined by a theory similar to that of the masses of a proton and electron; we cannot enter into the theory here.

Evidently we have here a definite starting-point for developing the theory of radiation on a quantum-relativistic basis. But our concern is with other matters, and we shall confine attention to a universe in its ground state. For our purpose the important conclusion is that the rest-energy attributed to the particles in our ordinary outlook is the energy of the limit state. Accordingly their rest-mass is m_2.

---

* The energy of a photon is the energy released when a threshold particle drops into the vacant energy level.

10—Since the energy levels near m_2c^2 are exceedingly close together, there is for ordinary purposes (but not for cosmological problems) an almost unlimited abundance of particles with rest-mass m_2. To the degree of approximation which represents the outlook of ordinary quantum theory, these particles do not "exclude" one another individually; for by a practically infinitesimal increment of energy a particle can reach a vacant level just above the limit energy. The usual applications of the exclusion principle, *e.g.*, in § 2, are to the relative motions resulting from the additional visible energy, and not to the energy representing the rest-mass.* Thus it is legitimate to assume (as in ordinary quantum theory) that each particle has an even *a priori* probability distribution throughout all space, which may be modified by special wave functions representing the visible energy and momentum of the particle. If the energy is precisely m_2c^2 so that the particle is at rest, its position is entirely indeterminable, *i.e.*, it is equally likely to be anywhere in space.

The total energy of the universe is $\mathrm{N}'\overline{m}c^2 = \frac{3}{5}\mathrm{N}'m_2c^2$ by (8.1). The deficit $\frac{2}{5}\mathrm{N}'m_2c^2$ is interpreted as gravitational potential energy, which is relatively large in a static (Einstein) universe.

Taking R to be the actual radius of space, the mass m_1 of a solitary particle in the K state is given by (7.82), and this has been regarded as the proper mass of a neutral particle. But in the current outlook the threshold mass m_2 is regarded as the proper mass. If we change the mass from m_1 to m_2, we must change the constant of gravitation from κ_1 to κ_2, so that

$$\kappa_1 m_1^2 = \kappa_2 m_2^2. \tag{10.1}$$

By a well-known formula for the Einstein universe the mass M of a static distribution occupying space of radius R is given by

$$\kappa \mathrm{M}/c^2 = \tfrac{1}{2}\pi \mathrm{R}. \tag{10.2}$$

Applying this firstly to the solitary particle and secondly to the mass $\frac{3}{5}$ $\mathrm{N}'m_2$ of the actual universe,

$$\kappa_1 m_1 = \tfrac{3}{5}\mathrm{N}'\kappa_2 m_2. \tag{10.3}$$

By (10.1) and (10.3)

$$m_2 = \tfrac{3}{5}\mathrm{N}'m_1. \tag{10.4}$$

---

* It should be remembered that an unconstrained uniform static distribution of density (Einstein world) is of order $\rho = 10^{-27}$ gm/cc. So that the exclusion effects at ordinary densities arise in the manipulation by ordinary quantum theory of the raw material investigated in the present theory.

The constant h will also be affected by the change of reckoning in such a way that

$$\mu_1 = m_1/h_1^2 = m_2/h_2^2 = \mu_2$$

and the current reckoning corresponds to $h = h_2$. Hence by (7.81) and (10.4)

$$m_2 = 3\,(\tfrac{3}{5}\mathrm{N}')/\mathrm{R}^2c^2\mu_2 \tag{10.51}$$

or

$$m_2 = \sqrt{\tfrac{3}{5}\mathrm{N}'}\,.\,h\sqrt{3}/2\pi c\mathrm{R}. \tag{10.52}$$

11—The factor 3 in (10.51) corresponds to the three components of the pressure. For a uniform probability distribution in n dimensions the corresponding factor would be n. Physically this corresponds to the law of equipartition of energy in statistical equilibrium; if the number of equivalent degrees of freedom is increased, the energy of the particle is increased in the same ratio. Similarly if a system, which has been moving as a single particle, is ionized into 4 independent particles the energy is multiplied by 4. Ordinarily these results are applied only to the energy additional to the rest-energy, but we here find that they apply also to the rest-energy.

When we replace the neutral particle by four charged particles which, being described by wave vectors, have additional relativistic rotations, the energy (and therefore the mass) is increased by the additional degrees of freedom. It is therefore desirable to define a standard mass m_0 corresponding to a definite number of degrees of freedom—to be used in much the same way as the energy of a monatomic molecule is used as standard in the theory of gases. We take m_0 to correspond to *two* degrees of freedom. The choice (explained in § 12) is guided by the fact that all observables must arise from double wave functions. Accordingly

$$m_0 = 2\,(\tfrac{3}{5}\mathrm{N}')/\mathrm{R}^2c^2\mu_0, \tag{11.1}$$

so that

$$m_0 = h\sqrt{\tfrac{3}{10}\mathrm{N}}/2\pi c\mathrm{R}, \tag{11.2}$$

where $\mathrm{N} = 4\mathrm{N}'$ = number of charged particles.

This is the mass which is referred to in my equation for the masses of the proton and electron,*

$$10m^2 - 136mm_0 + m_0^2 = 0. \tag{11.3}$$

Denoting the roots by m_p, m_e, we have

$$m_p + m_e = (136/10)\,m_0. \tag{11.4}$$

---

* 'Proc. Roy. Soc.,' A, vol. 143, p. 327 (1934). This identification is explained in § 12.

We have also by (10.2)

$$\tfrac{1}{2}\pi\mathrm{R} = \kappa\mathrm{M}/c^2 = (\kappa/c^2)\,\tfrac{1}{2}\mathrm{N}\,(m_p + m_e). \tag{11.5}$$

The constant κ in (11.5) necessarily differs from that which has occurred in previous formulae, *e.g.*, (10.3), because we have got more energy into space of the same curvature by allowing the particles additional degrees of freedom. We have not hitherto been concerned with the absolute value of κ; consequently (11.2) is unaffected by the change. The ordinary observational value is evidently the value in (11.5).

From (11.2) and (11.5)

$$\frac{\sqrt{\frac{3}{10}\mathrm{N}}}{\mathrm{R}} = \frac{2\pi c m_0}{h} \qquad \frac{\mathrm{N}}{\mathrm{R}} = \frac{\pi c^2}{\kappa\,(m_p + m_e)}. \tag{11.6}$$

Whence

$$\sqrt{\tfrac{10}{3}\mathrm{N}} = \frac{hc}{2\kappa m_0\,(m_r + m_e)} = \frac{136}{10}\,\frac{hc}{2\kappa\,(m_p + m_e)^2} \tag{11.7}$$

$$= \frac{136\cdot 137}{10}\,\frac{\pi}{\kappa}\left(\frac{e}{m_p + m_e}\right)^2, \tag{11.8}$$

since $hc/2\pi e^2 = 137$.

It is of interest to calculate N as accurately as possible. In order to obtain an accuracy better than 1% it is necessary to pay attention to the factor 136/137 pointed out by Bond, which, according to my theory, vitiates many of the so-called observational constants. I have therefore transformed the expression into the form (11.8), since the factor does not affect e/m_p.* Using Bond's determination from the latest available data,† and $\kappa = 6\cdot 664\,.\,10^{-8}$, we obtain

$$\mathrm{N} = 1\cdot 5727\,.\,10^{79}.$$

The uncertainty may be 1 or 2 parts in 1000.

Following Fürth's suggestion that the core of the number N is 2^{256}, we find that

$$\mathrm{N} = 135\cdot 82\,.\,2^{256}.$$

Since N is an integer by definition, no irrational or fractional factors can enter into its composition. The integers most directly connected with the theory are 2, 3, 4, 10, 16, 136, 137, 256. It therefore seems rather likely that the exact value is $\mathrm{N} = 136\,.\,2^{256}$.

---

* The "observational" values of e/m_e and m_e/m_p are both 136/137 of the "true" values. Consequently there is no error in e/m_p.

† 'Nature,' vol. 135, p. 825 (1935); $e/m_pc = 1\cdot 7703\,.\,10^7 \div 1847\cdot 6 = 9582$.

I regard this as a definitive determination of N, subject to the reservation that, since it was not originally intended to take account of factors of the order 136/137, these have not had the same scrutiny as the rest of the investigation.

The corresponding value of the Einstein radius R is $6{\cdot}17 \,.\, 10^{26}$ cm. The cosmical constant is $1/R^2$. The limiting speed of recession of the galaxies is 865 km per sec per megaparsec.

In (11.7) the factor 10/3 on the left is the 4-dimensional equivalent of the 16-dimensional factor 136/10 on the right. In other words, 10/3 has the same relation to Pauli matrices that 136/10 has to Dirac matrices. It would presumably have been possible to treat §§ 7–10 by Pauli matrices instead of by stereographic projection. The fact that the numerical factors in the answer have a natural interpretation diminishes our fear of slips in the deduction.

12*—It is incongruous that a scalar wave function with three degrees of freedom should represent four elementary particles. The explanation is that, by representing the waves in space of *fixed* radius R, we have suppressed all radial displacement and therefore all standing waves in the radial direction. But R, like every observable, is subject to the uncertainty principle, and has not a fixed value but a probability spread. In practice R is treated as fixed because, being the sole linear characteristic of the universe in its ground state, it is the standard to which all other lengths are ultimately referred; but this posterior use of it hides its intrinsic character. Before breaking up the neutral particle into four charged particles the radial degree of freedom should be restored. The scalar wave function with four degrees of freedom is then replaced by four vector wave functions as described in § 4.

We shall call the product of the rest-mass and the energy tensor the *quadratic energy tensor*. By (2.2) the probability distribution determines the quadratic energy tensor. Since m_0 has been standardized for two degrees of freedom the pressure corresponding to threshold neutral particle is $\frac{1}{2}m_0c^2$. Now let its probability distribution be modified so that the pressure is mc^2; and further let its scalar wave function be replaced by a vector with p symmetrical degrees of freedom. Then the quadratic pressure is m_0mc^2, and the quadratic energy is pm_0mc^2. In accordance with § 9 we try to represent this change as an addition to, instead of a modification of, the ground state. Determine q so that

$$pm_0mc^2 = qm^2c^2 + m_0{}^2c^2. \tag{12.1}$$

---

* This section is a brief outline, treated only so far as is necessary to show that m_0 is the same in (11.2) and (11.3).

Then the change is expressed as the addition of a particle of mass m and energy qmc^2 to a background particle of mass m_0 and energy m_0c^2, *i.e.*, the addition of a particle of mass m represented by a vector wave function with q degrees of freedom to a particle of mass m_0 represented by a scalar function with one degree of freedom, the masses now being standardized for one degree of freedom.

I have shown (*loc. cit.*) that if the new particle is an elementary particle representable by a simple wave vector, $p = 136$ and $q = 10$. Then (12.1) becomes the equation (11.3), giving two possible masses for an added elementary particle. The two values are therefore identified with m_p, m_e.

If we treat the unmodified neutral particle in the same way, changing neither its probability distribution nor the scalar character of its wave function, the equation corresponding to (12.1) is

$$2m_0^2c^2 = m^2c^2 + m_0^2c^2, \tag{12.2}$$

since its energy (for four degrees of freedom) is $2m_0c^2$. Hence $m = m_0$. We had (12.2) in view when we adopted two degrees of freedom as the standard in defining m_0. The result is that the masses m_p, m_e, m_0 are all treated on a uniform plan as relating to added particles. These masses are now standardized for one degree of freedom and therefore correspond with the reckoning of elementary quantum theory which represents the particles by plane progressive waves displaceable only along the wave normal. Further, since each is now represented as an addition to a background particle with one degree of freedom, the energy of the original standing wave with four degrees of freedom corresponds to four such background particles.

It may help to an understanding of the above investigation if we give a more elementary illustration of the factor 2 which is the main source of complication. The energy of a charge e is $eV = \Sigma ee'/r$ if it is treated as being in a field of fixed potential. But if we calculate in this way the energy of all the charges in a system we obtain twice the proper amount. It is to avoid a corresponding error in the present discussion that we initially standardize the mass for two degrees of freedom.

13—The results here obtained lead to many more developments than can be treated in a single paper. I will add one further comment. The constants of nature here found by considering a static Einstein universe will apply to the actual universe, notwithstanding that it is far from the static state. For just as in an experimental determination the matter is artificially arranged to simplify the experiment, so in a theoretical determination the matter may be artificially arranged to simplify the calculation. The only restriction is that the arrangement must be one that could be brought about in the actual universe without violating the equations of the theory, *e.g.*, we must not create or destroy particles.

The usual theory of the expanding universe (neglecting radiation) recognizes three possibilities according as the mass M is greater than, equal to, or less than the mass M_e of an Einstein universe. It would be illegitimate to apply our results to the actual universe unless $M = M_e$. But in the present investigation we see that $M = M_e$ of necessity, and that there is no opportunity for the other two alternatives to arise. By the exclusion principle a system of N particles must necessarily have a ground state, and therefore a static configuration is always possible; the contingency contemplated in ordinary relativity theory that there may be no static (Einstein) state can therefore be discarded. A similar argument applies when the amount of radiation in the universe is not neglected.

The following is an analytical view of this conclusion. By the ordinary formula $M_e = \frac{1}{2}\pi\lambda^{-\frac{1}{2}}$. It has therefore generally been supposed that M_e is determined by a constant λ fixed independently of the content of the universe. But primarily λ is the arbitrary constant in (2.1); and although later it is found desirable to fix it by convention (7.62), it is fixed with reference to the actual content of the universe, so that $M = \frac{1}{2}\pi\lambda^{-\frac{1}{2}}$. The result is that λ is a function of N, and not a predetermined constant.

The Preparation of Oxygen by Electrolysis of Baryta

By Eric M. Stoddart, B.Sc., Ph.D., Assistant Lecturer in Chemistry,
University College, Swansea

(*Communicated by F. G. Donnan, F.R.S.—Received May* 11, 1935)

Morley, in his classical researches, pointed out that various impurities were present in the gases produced by electrolysis of sulphuric acid and potassium hydroxide solutions. Using the latter method, Lord Rayleigh,* by addition of a small quantity of baryta, attempted to minimize the amount of hydrocarbons present in the hydrogen as impurity, derived from the carbonate present in the alkali. Baker† pointed out that the action of baryta on potassium carbonate is a reversible one, particularly in the presence of such a large excess of potash, and concluded that Rayleigh's precaution might not be particularly effective. Baker recommended the electrolysis of pure barium hydroxide solution, in which barium carbonate is insoluble, and which consequently must give hydrogen free from hydrocarbons. He also noted that, with the use of pure baryta solution, the oxygen is apparently free from ozone and hydrogen peroxide. It is to be noted that Baker gave no details concerning the shape and material of his electrodes, or of the currents employed in the electrolysis.

The method of electrolysing baryta, using nickel electrodes, has become a standard method of preparing small quantities of pure oxygen. The present author has employed the apparatus shown in fig. 1 for this purpose during the past few months. The apparatus consists of two vessels, A and B, constructed from 1-inch internal diameter tubing, carrying nickel electrodes, the upper ends of which are welded to platinum wires sealed through the glass, the seals being sufficiently strong to allow currents up to 2 amperes to pass without cracking the glass. The use of nickel is justified owing to its low hydrogen and oxygen over-voltages. The electrodes are constructed from 18 S.W.G. pure nickel wire, 3 feet of which is coiled up as shown. These electrodes have a fairly large surface area—approximately 34 sq cm. The baryta solution was saturated at room temperature and a layer of pure barium hydroxide was allowed to settle on the bottom of the outer jar. The electrolysis was carried out, using a current of 1 ampere, which kept the temperature of the solution at about 60° C. The oxygen was used in experiments, and was found at first to be highly satisfactory.

---

* 'Proc. Roy. Soc.,' A, vol. 45, p. 425 (1889).

† 'J. Chem. Soc.,' p. 400 (1902).

As time went on, however, the gases were no longer evolved in the ratio $H_2/O_2 = 2/1$. The ratio approximated to 1/1, and it was found that the oxygen contained large concentrations of hydrogen—in fact, the gases were highly explosive. Mellor* comments on the fact that there have been several explosions from the use of electrolytic oxygen, "owing to the hydrogen getting mixed therewith." The apparatus used in the present work was designed to avoid any possibility of such an occurrence, and it is clear that the mixed gases were actually evolved together at the anode.

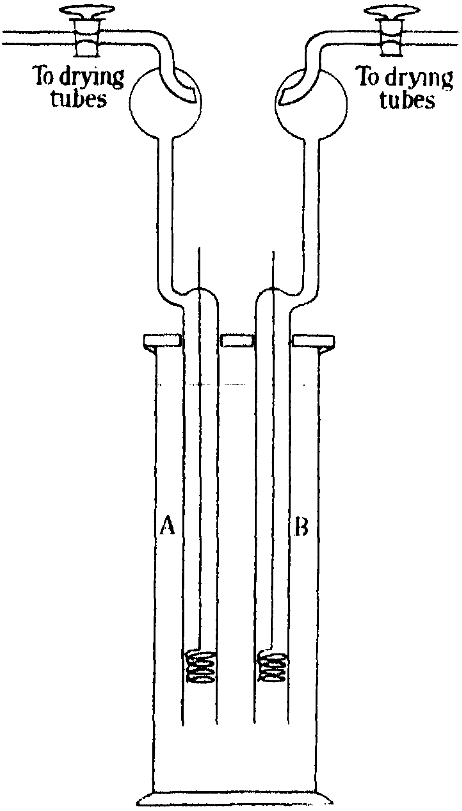


FIG. 1.

The current density in these preparations was only 30 milliamperes per sq cm, which is comparatively low. When the density was reduced to

---

* "Comprehensive Treatise of Inorganic Chemistry," vol. 1, p. 356 (1922).

10 milliamperes per sq cm the oxygen evolved was free from hydrogen, but when the current was restored to its initial value an explosive mixture was again evolved.

The following treatment was found effective in preventing the liberation of the hydrogen at the anode. A reversed current of $\frac{1}{4}$ ampere was passed through the cell for 12 hours, so causing the liberation of hydrogen at the electrode designed as anode. The current was reduced to a few milliamperes and reversed, thus liberating oxygen at the anode, which became coated with a thin adherent layer of the black hydrated oxide of nickel. A mild electrolysis was allowed to proceed for about 6 hours, after which it was found that the cell would take currents up to 2 or 3 amperes, evolving only oxygen at the anode. In this way, the cell was used for about 6 months, but finally the mixed gases were again evolved, causing a rather disastrous explosion. It was noted that the mixed gases evolved had *stripped the oxide film from the electrode. On reconditioning* the anode, the cell was fit for use again, but after a few days the mixed gases were again liberated. Repetition of the reconditioning process finally led to a nickel anode which refused to become coated with an oxide layer, and always evolved the mixed gases, except at extremely low-current densities—about 1 to 2 milliamperes per sq cm.

The only explanation for these observations seems to be that electrolysis at the anode takes place as follows:

$$OH^- + OH^- = OH + OH + 2e$$

$$OH + OH = H_2O + \tfrac{1}{2}O_2 \quad \text{or } \tfrac{1}{2} \text{ volume} \qquad \text{(I)}$$

$$OH + OH = H_2 + O_2 \quad \text{or 2 volumes.} \qquad \text{(II)}$$

Under normal conditions of electrolysis, process (I) is the main reaction, but conditions can favour process (II) to some extent. It is a simple matter to show that if the evolved gases are in the ratio 1/1, then 2/3 of the disappearance of the hydroxyl is due to (I) and 1/3 due to (II). Consequently, the gases would be expected to contain about 1/3 hydrogen. Analysis by explosion showed a 30% hydrogen content.

The fact that a layer of oxide on the anode can prevent the liberation of hydrogen is rather difficult to account for. The layer is presumably thin, and would not be expected to remove the hydrogen by direct oxidation. It is possible that the actual anode process is always according to (II), followed by the reaction $H_2 + O_2 = H_2O + \frac{1}{2}O_2$. This mechanism would account for the fact that temporary losses occur at the beginning of such an electrolysis, caused by the recombination of hydrogen and oxygen dissolved in the electrolyte, the dissolved hydrogen uniting with the evolved oxygen on the nickel surface.*

If the nickel surface lost its power to bring about this combination of the gases, or, in other words, if its surface became " poisoned " towards this reaction, then the mixed gases would be evolved. That this seems to be the true explanation is borne out by the fact that after a time the surface cannot be restored to any condition which prevents this liberation of mixed gases, a phenomenon indicating a poisoning effect.

The preparation of hydroxyl in gases at low pressures has been the subject of many researches. Lavin and Stewart† reported that such hydroxyl disappeared according to the reaction:

$$OH + OH = H_2O_2,$$

but their observations could not be repeated by Bonhoeffer and Pearson‡ or by the present author.§ Since the hydroxyl in the present experiments is produced under very different conditions to those in an electrical discharge, it was of interest to determine whether hydrogen peroxide could be detected in the liquid round the anode.

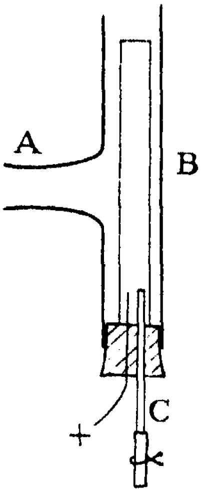


FIG. 2.

An ordinary gas voltameter was employed for the experiment, the anode having the appearance as shown in fig. 2. The tube A was the bridge between the electrode chambers. The anode consisted of ½-inch 22 S.W.G. pure nickel wire; this produces a high-current density which seems to favour evolution of mixed gases. The electrode was surrounded by the glass sheath B, which served to direct the gases up past the bridge tube A, thereby preventing mechanical mixing of the gases. The anode liquor was removed via tube C and examined for the presence of peroxide. This was done by precipitating the barium by addition of excess dilute sulphuric acid and testing the clear liquid by the titanium reaction. Peroxides were always absent.

It was of interest to note that using such a small electrode, and passing 1 ampere through the cell, after about 6 hours gases were evolved containing as much as 20% hydrogen. This could be prevented by an oxide coating as before, but, owing to the rapid evolution of gas, the oxide was quickly stripped off and the mixed gases again evolved.

---

* *See* Mellor, *op. cit.*, vol. 1, p. 276 (1922).

† 'Proc. Nat. Acad. Sci. Wash.,' vol. 15, p. 829 (1929).

‡ 'Z. phys. Chem.,' B, vol. 14, p. 1 (1931).

§ 'Phil. Mag.,' vol. 18, p. 409 (1934).

SUMMARY

It has been shown that electrolysis of baryta, using nickel electrodes, does not always give pure oxygen at the anode. After some time, hydrogen makes its appearance along with the oxygen, probably owing to the reaction:

$$OH + OH = H_2 + O_2.$$

Liberation of hydrogen at the anode can be prevented by coating the electrode with a layer of the black oxide of nickel.

Aspects of the Current System Producing Magnetic Disturbance

By J. M. STAGG, M.A., B.Sc.

(*Communicated by G. C. Simpson, F.R.S.—Received May* 13—*Revised June* 17, 1935)

1—If N, E, and Z (vertical) are the three rectangular components of the earth's magnetic field, a study of the behaviour of the changes ΔN, ΔE, and ΔZ produced in them by a perturbing field gives information about the corresponding changes in this field and therefore in the overhead current system to which the field is due. In this enquiry ΔN, ΔE, ΔZ will be regarded as departures for 60-minute intervals from an undisturbed condition which is taken to be the corresponding departure on quiet days. It will be understood that this procedure assumes that the effect of disturbance on the earth's field is to overlay quiet conditions by an additional system of forces. Average diurnal variations on both quiet, q, and disturbed, d, days were therefore formed for N, E, and Z, starting from the published hourly values of the primary horizontal elements H and D where necessary. With the two sets of variations converted into departures from their respective means, difference departures $\Delta d - \Delta q$ for N, E, and Z were formed hour-by-hour to give 24 disturbance field component vectors. This was done for the seasons separately and the year as a whole for each of the observatories in Table I.

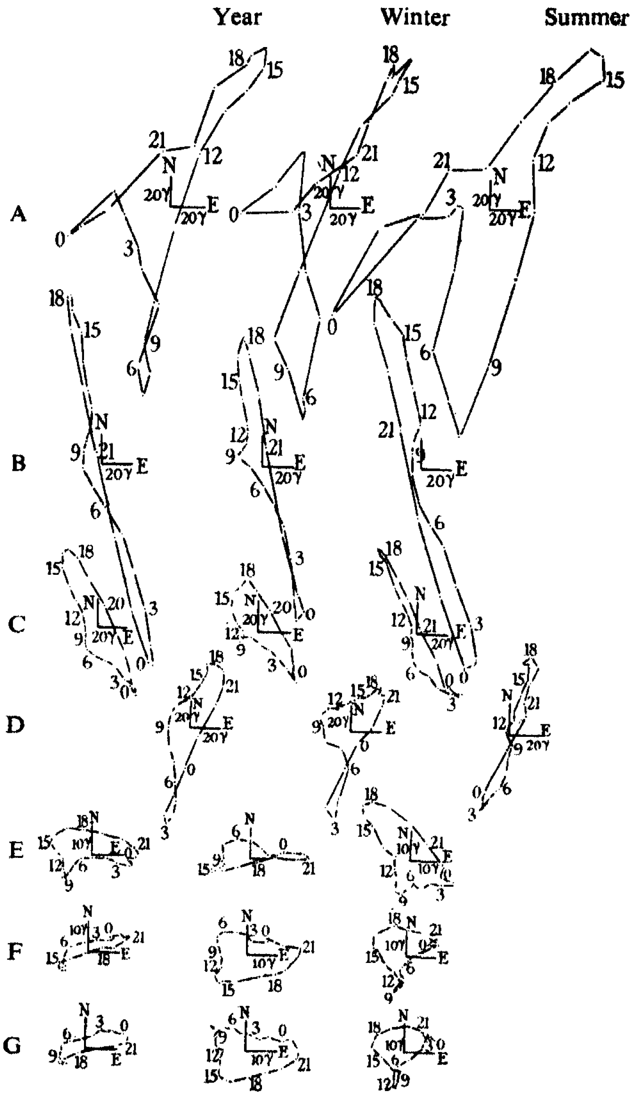
In discussing the average features of regular disturbance there are advantages in using as extensive series of data as possible. Even when the five internationally selected *d* and *q* days per month are used, as in the present enquiry, exceptional isolated disturbances may obscure normal behaviour. This is especially true when the study concerns stations representing transition zones in the phenomena of disturbance. The inclusion of Fort Rae, the British Polar Year station in N.W. Canada, will, however, be seen to be fully warranted, not because of the scale of disturbance there but because of its instructive position on the polar side of the line over which the current system producing the main features of the regular daily variation of disturbance is mainly concentrated.

TABLE I

Observatory	Geographical		Magnetic latitude	Years used	Elements available	Mean declination
	Latitude	Longitude				
	°	°	°			°
Sodankylä	67·4	26·6	63·8	1914–23	H, D, Z	0·9 E
Lerwick	60·1	358·8	62·6	1926–33	H, D, Z	14·3 W
Sitka	57·0	224·7	60·0	1915–23	H, D, Z	30·4 E
Eskdalemuir ..	55·3	356·8	58·5	1911–22	N, E, Z	17·3 W
De Bilt	52·1	5·2	53·8	1911–22	H, D, Z	11·9 W
Seddin	52·3	13·0	52·4	1914–24	N, E, Z	7·7 W
Fort Rae	62·8	116·1	69·1	1932–33	H, D, Z	37·5 E

2—Figs. 1 and 2 show for the whole period used and for the winter (January, February, November, December) and summer (May to August) months the disturbance vector diagrams in the horizontal (N, E) and vertical meridian (N, Z) planes. Throughout these diagrams the hours as numbered are the nearest local whole hours. In both figures the scale for the three most southerly observatories is double that of the remaining four. The great increase in the scale of the regular disturbance vector accompanying the 17° increase of (magnetic) latitude between Seddin and Fort Rae can best be commented on numerically when the resultant whole-field vector is discussed, § 6. Other less obvious features may also be more suitably deferred till the co-ordinates of this resultant vector are available, §§ 7, 8. Here examination is confined to those aspects to which the diagrams themselves give the readiest clue.

The common characteristic differentiating the N, E vector diagrams, fig. 1, for Seddin, De Bilt, and Eskdalemuir from those of more northerly stations is the tendency of the former to be elongated transverse to the magnetic meridian (*see* values of average declination in Table I). But already at Eskdalemuir in summer the diagram is elongated into the N.W. quadrant. At the latitude of Sodankylä the vector transverse to



A, Fort Rae; B, Sodankylä; C, Lerwick; D, Sitka; E, Eskdalemuir; F, De Bilt; G, Seddin.

FIG. 1—Vector diagrams in horizontal plane. Scale for Eskdalemuir, De Bilt, and Seddin double that for other stations.

the meridian practically vanishes. The diagram for Fort Rae is highly irregular but on the whole lies along the meridian. As was first pointed out by Lüdeling,* roughly oval or circular diagrams may be expected at stations in still higher latitudes.

As the Eskdalemuir diagram in summer tends to the elongated form of more northerly stations, so the diagrams for Lerwick and Sitka tend in winter to the form characteristic of the lower latitude stations where the elongation is transverse to the meridian. The Fort Rae diagrams, both in winter and summer, are more highly irregular than may be attributed solely to the short period covered by the data. In winter the diagram is more elongated as well as more complex than in summer.

TABLE II

Observatory	Magnetic latitude °	Year	Winter	Summer	Zone
Gjöahavn	78	*cc*			pz_1
Fort Rae	69	*ac*	—	*ac*	z_1
Sodankylä	64	—	*cc* *ac*	*ac*	z_2
Lerwick	63	*cc*	*cc*	*cc*	z_2
Sitka	60	*cc*	*cc* *ac*	*cc* *ac*	z_2
Eskdalemuir	58	*cc*	*ac*	*cc*	z_2
De Bilt	54	*ac*	*ac*	*cc* *ac*	ez_2
Seddin	52	*ac*	*ac*	*cc* *ac*	ez_2

3—Intimately connected with the form of the horizontal disturbance vector diagrams, other characteristics shown by the directions in which the diagrams are described are recognizable in fig. 1. These are summarized in Table II in which *cc* denotes rotation with, and *ac* against, the clock. If the vector clearly rotates in both directions, that followed in the greater number of hours is given first; a dash indicates a complexity of rotation. The mechanism producing these changes in rotation direction and form of the vector diagrams presently being considered need not be discussed in detail; they are used only as indices of the zonal arrangement of the disturbance field. It is enough to say that they are to be explained solely by the relative magnitudes and phases of the disturbing vectors.

---

* 'Terr. Magn. and Atmos. Electricity,' vol. 4, p. 254 (1899).

It should be noted that the *ac* direction for Fort Rae is also that of other similarly situated stations examined by Lüdeling, *loc. cit.*, and the opposite direction for higher latitudes exemplified in Lüdeling's enquiry by Kingua Fjord is confirmed by Graarud and Russeltvedt* for Gjöahavn.

The grouping of stations suggested by Table II is the same as that of § 2. At Sodankylä, situated, as will be seen presently, just to the south of the line of greatest current concentration, the rotation direction is indefinite. Southward, to a limit about Eskdalemuir, the rotation is *cc*; northward, to about the latitude of Fort Rae, it is *ac*. Beyond these limits again the rotation changes to *ac* below Eskdalemuir and to *cc* above Fort Rae. The various zones suggested by Table II may conveniently be referred to by letters:—z_1 is the zone between Sodankylä and Fort Rae, z_2 between Sodankylä and Eskdalemuir; pz_1 is the region on the polar side of z_1 and ez_2 on the equatorial side of z_2. The line just to the north of Sodankylä between z_1 and z_2 will be referred to as the current line.

Table II shows that the limits between z_1 and pz_1 and between z_2 and ez_2 vary with season. In summer the northern limit of ez_2 moves south to include De Bilt and Seddin; in winter it retreats so that the *ac* direction appropriate to ez_2 becomes the characteristic of Eskdalemuir and even Sitka.

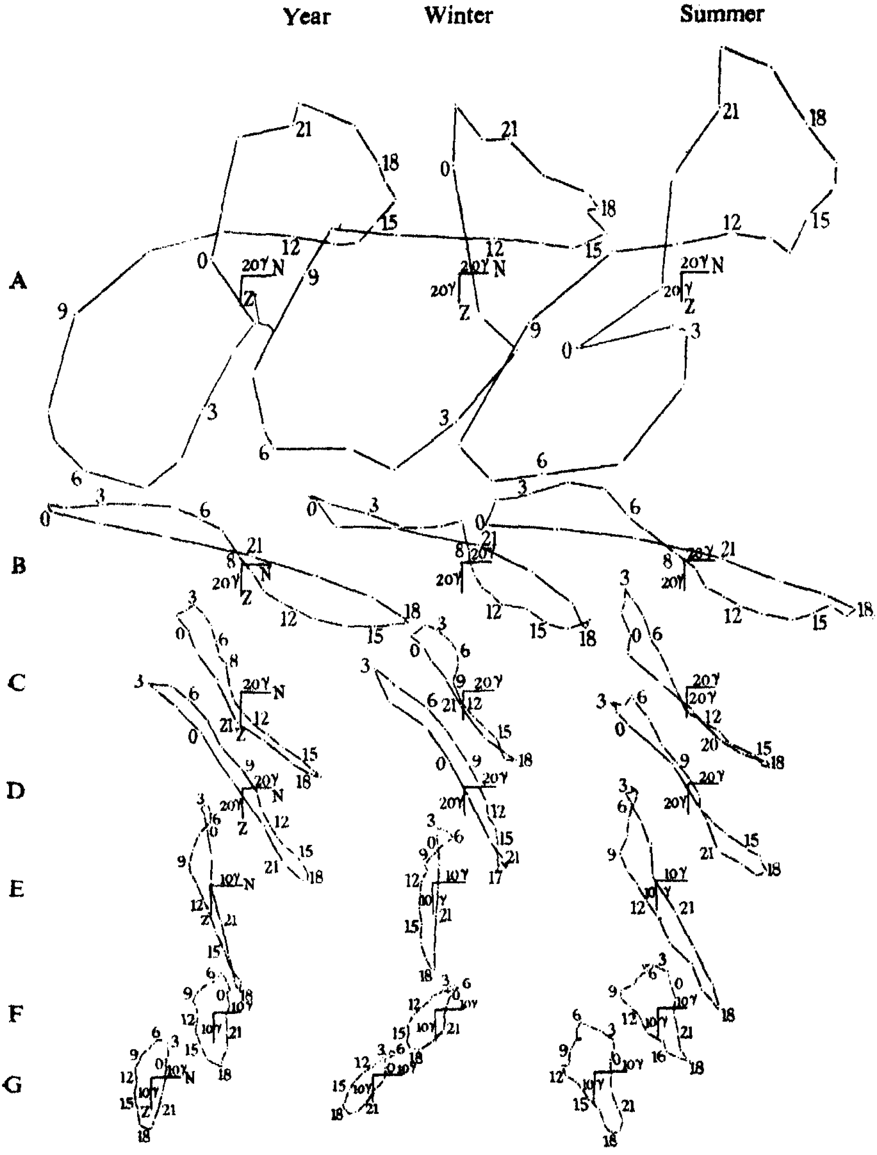
North of the current line the corresponding seasonal changes are less unambiguous. But considered along with the winter and summer shapes of the vector diagrams, § 2, the rotation direction at Fort Rae may be interpreted in this way. A rounded form and *cc* rotation are the characteristics of the vector diagram for the horizontal disturbance field within pz_1. At Fort Rae, though the transverse field is more highly developed in winter than in summer so that the diagram is broader in this latter season, the vector rotation then is consistently *ac*. Hence z_1 is probably affected by the same southward expansion in summer as the zones on the equatorial side of the current line.

Lerwick is the only one of the seven stations whose direction of rotation remains constant in all seasons.

Some of the seasonal changes in the behaviour of the disturbing vector in the horizontal plane described above can be ascribed either to horizontal or vertical modifications in the space distribution of the current system. Increased height in summer with consequent extended horizon would alone adequately account for zonal movements in localities south of the current line. Such a summer increase in height has indeed been

---

* 'Geofys. Publikasjoner,' vol. 3, p. 12 (1925).



A, Fort Rae; B, Sodankylä; C, Lerwick; D, Sitka; E, Eskdalemuir; F, De Bilt; G, Seddin.

Fig. 2—Meridian plane components. Scale for Eskdalemuir, De Bilt, and Seddin double that for other stations.

found by Goldie,* using a wholly different mode of investigation. But, as will be seen more clearly in § 10, the real alternative to a southward movement in summer of all zones is probably an increased height in that part of the current system controlling localities in z_2 and ez_2 and a decreased height in z_1.

4—The disturbance vectors in the meridian plane, fig. 2, will now be considered, restricting attention at first to those for the year in the left-hand side of the figure. Noting that in these (as in the separate seasonal diagrams) direction to north is towards the right of the figure, it is clear that at Fort Rae the main slope of the vector diagram away from the vertical is to the north, while for all the other stations, except De Bilt and Seddin, the slope is to the south. Further, this slope is greatest at Sodankylä and becomes increasingly less with decreasing latitude until at De Bilt and Seddin the diagram is vertical or slightly inclined to the north. If we suppose all seven diagrams simplified so that each may be replaced by a pair of vectors acting, one through the hours of maximum vector directed downward, the other through the hours of maximum upward vector, we may infer that, if the current producing these diagrams is simple and concentrated, it is situated overhead between the latitudes of Sodanyklä and Fort Rae and probably nearer Sodankylä than Fort Rae. We may also infer that Eskdalemuir is just to the north of (*i.e.*, within) the horizon of the concentrated current, De Bilt is just on it, and Seddin beyond it, so that at this last station the vector in the vertical plane appears to be related to a current below ground level. These are the average effects of all seasons.

The separate winter and summer meridian diagrams point to seasonal changes of the same nature as those deduced from the horizontal components. At Eskdalemuir, for example, the winter figure is vertical; in summer it is inclined southward at an angle of about 25°. At De Bilt and Seddin the winter figures slope to the north; in summer both are tilted slightly southward. For those three southern stations these changes imply that in winter the southern limit of northward inclination retreats northward to the latitude of Eskdalemuir; in summer it advances to include De Bilt and Seddin. For the more northern stations the corresponding seasonal changes are not so conspicuous largely because of the complex form of the figures, especially that of Fort Rae. But at Sodankylä there is evidence that the vector is more nearly horizontal in summer than in winter at the time when the vector is greatest.

5—Other features of interest in fig. 2 relating to the approximate times of greatest upward and downward directed vector and times of transition are summarized in Table III. At about 18 hr, when the meridian vector is greatest upward at Fort Rae it is greatest downward at the other six stations south of the current line. The immediate inference is that the direction of current flow producing the maximum vector at this and neighbouring evening hours is from west to east; in the early morning hours centred about 4–5 hr the current is at its maximum in the reverse direction. The local time of maximum west-east current is practically the same at all stations; the morning time of maximum east-west current is 2 or 3 hours later in z_1 and ez_2 than in z_2.

---

* 'Trans. Roy. Soc. Edin.,' vol. 62, p. 161 (1931).

The times of reversal of vector—and therefore current—direction deduced from fig. 2 are also shown in Table III. For the stations in z_2 and ez_2 21 hr is general for the reversal from west-east to east-west current; at Fort Rae it is 3 or 4 hours later. The opposite change occurs at all stations in an interval of 3 hours before local noon. The evening reversal in direction of disturbing current affecting Lerwick and Eskdalemuir has been noted by Goldie, *loc. cit.*, p. 157.

TABLE III

	Local time of maximum disturbance current		Local time of reversal of disturbance current	
	W.-E. hr	E.-W. hr	W.-E. to E.-W. hr	E.-W. to W.-E. hr
Fort Rae	19	6	0–1	9–10
Sodankylä	17	5	20–21	8–9
Lerwick	17	3	21–22	9
Sitka	18	3	22–23	10
Eskdalemuir	18	3	22	11
De Bilt	18	5	22	12
Seddin	18	6	23	12

6—Further detailed examination of the characteristics of the $(d-q)$ vectors is assisted by converting the rectangular components of each hourly vector into polar co-ordinates R, θ, φ, where R is the mean hourly resultant vector, θ is the inclination of R to the vertical measured from the nadir, and φ is the angle, measured from north through east, which the vertical plane containing R makes with the vertical plane through the geographical meridian. The values of R, θ, and φ for each hour are determined by

$$R^2 = \Delta N^2 + \Delta E^2 + \Delta Z^2 = \rho^2 + \Delta Z^2$$
$$\theta = \tan^{-1} \Delta E / \Delta N$$

and

$$\phi = \tan^{-1} \rho / \Delta Z,$$

ρ being the resultant vector in the horizontal plane. Table IV gives the values of R, θ, and ϕ for each hour of the day for the year as a whole for the seven stations.

The diurnal variation of the resultant vector R is shown graphically in fig. 3. At each station R has a primary and secondary maximum each day, the one in the evening and the other in the early morning hours.

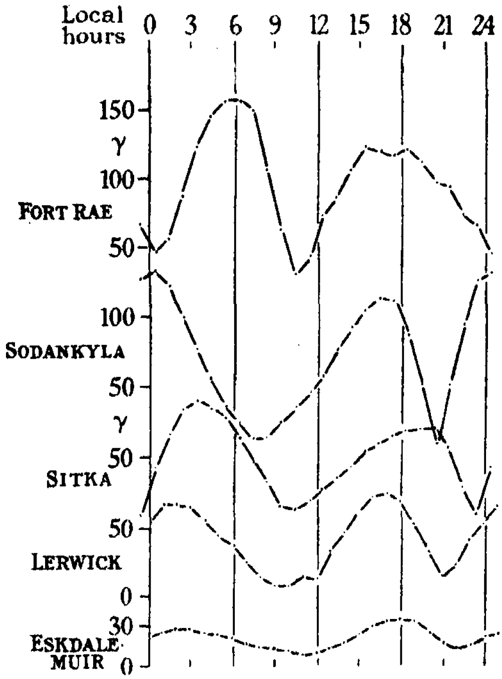


FIG. 3—Diurnal variation of total field resultant vector.

It is to be noted that at Sitka, as at Sodankylä and Fort Rae, the primary maximum is in the morning. R at Fort Rae reaches a maximum a 12-hourly intervals; at the stations on the equatorial side of the current zone the evening maximum occurs 16 hours after the morning maximum. This implies that the west-east current of the evening at these latter stations decreases to a minimum, reverses and rises to a maximum from the east in half the time taken for the opposite change, whereas at Fort Rae the current variations are almost purely semi-diurnal.

Table V gives the average daily values of R; R′ is the corresponding value corrected for the magnetic activity of the years from which R is

TABLE IV—HOURLY MAGNITUDES AND ORIENTATION OF DISTURBANCE WHOLE-FIELD VECTOR

Local hour	Sodankylä			Lerwick			Sitka			Eskdalemuir			De Bilt			Seddin			Fort Rae		
	R	θ	φ	R	θ	φ	R	θ	φ	R	θ	φ	R	θ	φ	R	θ	φ	R	θ	φ
	γ	°	°	γ	°	°	γ	°	°	γ	°	°	γ	°	°	γ	°	°	γ	°	°
1	134	106	164	69	135	147	44	143	187	25	149	98	13	135	61	12	124	57	49	58	285
2	123	107	163	68	141	145	67	141	192	27	152	105	13	145	59	12	138	50	57	50	260
3	100	112	161	66	147	149	84	138	195	27	161	104	12	159	34	13	146	43	89	47	220
4	78	119	161	55	152	154	90	137	193	25	169	97	13	167	346	13	157	35	123	47	204
5	56	130	162	45	155	169	85	140	190	23	179	126	13	157	294	13	165	342	146	48	187
6	37	138	173	39	144	187	79	144	191	20	170	258	13	151	282	13	156	295	157	53	191
7	25	147	207	26	146	206	65	147	194	17	155	236	12	140	255	13	145	278	157	57	187
8	13	118	266	16	131	216	47	150	204	15	140	231	11	132	245	11	139	260	150	61	185
9	14	68	307	10	98	227	32	146	226	13	127	228	11	121	233	11	128	249	109	78	188
10	24	49	326	10	51	244	18	132	270	11	107	231	11	111	232	10	118	243	62	104	189
11	36	49	343	16	28	275	15	84	306	9	89	240	8	100	229	9	106	234	31	132	165
12	47	52	347	13	32	312	20	50	327	10	63	250	8	80	233	8	93	232	45	118	28
13	59	55	348	35	39	322	29	37	351	13	46	266	10	67	233	8	89	232	71	105	29
14	77	60	348	48	42	326	38	36	6	18	38	276	12	54	238	10	62	233	87	102	33
15	93	65	349	61	46	330	45	38	13	24	33	288	14	43	249	12	50	239	107	106	33
16	106	69	348	72	46	334	55	40	15	30	27	298	16	23	247	15	40	243	124	110	30
17	114	72	348	75	45	338	60	40	15	33	21	316	16	2	255	16	26	240	120	116	27
18	111	72	347	70	42	342	67	38	13	34	15	343	17	18	72	17	9	237	119	120	25
19	89	77	376	53	37	350	70	37	19	33	15	27	17	30	77	17	16	88	121	128	20
20	51	82	345	34	27	15	72	36	26	25	26	63	17	49	72	17	34	87	111	134	20
21	10	118	333	16	51	148	71	36	29	18	50	79	15	64	71	16	51	80	99	133	351
22	51	111	168	25	111	129	57	34	36	14	91	89	13	82	70	15	70	74	95	132	266
23	96	106	167	44	125	144	30	42	46	16	123	91	12	103	63	14	90	70	72	120	258
24	129	105	165	55	134	147	10	131	138	21	146	98	11	128	58	13	107	66	67	98	253

TABLE V

	Fort Rae	Sodankylä	Lerwick	Sitka	Eskdalemuir	De Bilt	Seddin
R (γ)	99	70	43	52	21	13	13
u_1 activity measure	(35)	56·4	55·9	59·6	50·1	50·1	56·4
Factor for reference to Seddin standard	1·61	1·00	1·01	0·95	1·13	1·13	1·00
R′ (γ)	159	70	43	49	24	15	13
ρ (γ)	69	63	28	32	10	8	8
ρ' (γ)	111	63	28	30	11	9	8

derived. To obtain R′ mean values of the u_1 measure of activity developed by Bartels* were formed for the various sets of years used. Then assuming that the scale of disturbance in the latitudes concerned waxes and wanes proportionally with u_1, the vectors for each station were referred to the standard of disturbance in the years covered by the Seddin data by multiplication by the factors given in the third line of Table V.

From Seddin to Sodankylä (except between Sitka and Lerwick) R′ increases with approach to the current zone. If the purely qualitative inferences in earlier paragraphs regarding the average position of the disturbance current system related to a concentrated overhead current between Sodankylä and Fort Rae, the field strength would fall off symmetrically to north and south of the current line, so that R′ at Fort Rae should be comparable with R′ at Sitka or Lerwick. That the average regular disturbance at Fort Rae is over three times as great indicates that other complicating factors enter.

These complications may arise first from assuming that the current distribution is symmetrical about the earth's magnetic axis pole. The anomalous position of Sitka in R′, as well as in the other diurnal characteristics of R shown in fig. 3, and still others to be noted later, give additional ground for believing that this assumption is not justified. At the same time any asymmetry would by itself be unlikely to be the whole explanation of the high R′ at Fort Rae. Secondly, the complication may also arise from a complex distribution of current flow in the overhead layers between Sodankylä and Fort Rae. If the distribution were in the form of a sheet extending nearly to the latitude of Sodankylä in the south and less closely to the latitude of Fort Rae in the north, but at less than half the height along this northern periphery of the sheet, the main features of R would be accounted for. In this connexion it is worth notice that the values of the resultant vector, ρ, in the horizontal plane (also given in Table V) show that, though the great increase in R between Sodankylä and Fort Rae is shared by both the horizontal and vertical disturbing vectors, it is the vector in the vertical plane which makes the greatest contribution to the increase.

Defining the day hours as 7 hr to 18 hr inclusive and the remaining twelve as night hours, Table VI, *a*, shows the mean day and night values of R for the year as a whole and for the four winter and four summer months separately; Table VI, *b*, does the same for ρ. As might perhaps have been deduced from figs. 1 and 2, Table VI, *a*, shows that on the average of

---

* 'Terr. Magn. and Atmos. Electricity,' vol. 37, p. 1 (1932), and vol. 39, p. 1 (1934).

Table VI—Mean Values of R and ρ during "Day" and "Night" Hours

		Fort Rae	Sodankylä	Lerwick	Sitka	Eskdalemuir	De Bilt	Seddin
			a—R					
Year	24 hours	98·7	69·6	42·5	52·0	20·8	12·8	12·8
	Day	98·5	59·8	37·6	40·9	19·0	12·2	11·9
	Night	98·8	79·5	47·5	63·2	22·7	13·4	13·7
	Night/day	1·00	1·33	1·26	1·55	1·19	1·10	1·15
Winter..............	24 hours	95·8	54·0	32·7	45·9	17·2		
	Day	93·4	45·5	28·6	35·3	15·9		
	Night	98·2	62·5	36·9	56·4	18·6		
	Night/day	1·05	1·37	1·29	1·60	1·17		
Summer	24 hours	96·5	71·6	44·4	46·6	24·3		
	Day	103·5	59·0	38·7	36·0	23·1		
	Night	96·1	84·1	50·2	57·2	25·5		
	Night/day	0·93	1·43	1·30	1·59	1·10		
			b—ρ					
Year	24 hours	68·8	63·3	27·8	32·3	10·1	8·1	8·3
	Day	86·0	52·7	26·1	25·2	10·4	7·7	7·6
	Night	51·6	73·8	29·5	39·4	9·7	8·5	9·1
	Night/day	0·60	1·40	1·13	1·56	0·93		
Winter..............	24 hours	68·3	46·6	20·1	25·7	9·5		
	Day	85·9	37·0	18·6	19·5	9·1		
	Night	50·7	56·2	21·5	31·8	10·0		
	Night/day	0·59	1·52	1·16	1·63	1·10		
Summer	24 hours	68·1	65·2	29·4	28·9	12·1		
	Day	88·3	52·3	26·5	21·4	12·3		
	Night	47·8	78·1	32·3	36·4	11·9		
	Night/day	0·54	1·49	1·22	1·70	0·97		

the year R at Fort Rae, zone z_1, and at stations in ez_2 is almost equally developed in the light and dark hours; only in z_2, comprising stations immediately to the south of the current line, is disturbance generally greater in the dark hours. A more unexpected feature is that the summer and winter values of the night/day ratio should be so similar within each station irrespective of its geographic latitude. Part *b* of Table VI shows that this feature is equally true of ρ; it is therefore true of the remaining component in the vertical plane. But whereas the night/day ratios of disturbance are essentially similar for R and ρ at most stations (though differing from station to station in a way suggesting more than a merely zonal influence), the value of this ratio for ρ at Fort Rae is little more than half that for R.

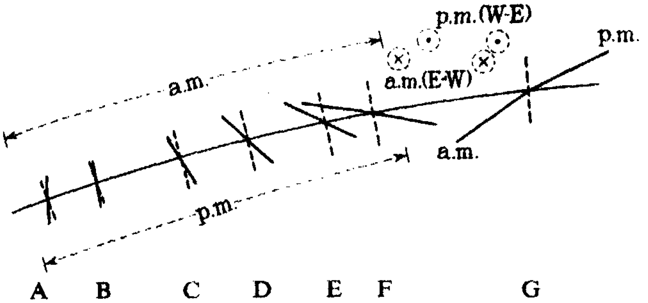
7—Even with the additional set of vector diagrams analogous to those of figs. 1 and 2, drawn to represent the daily behaviour of the disturbance vector in the prime vertical (E, Z) plane, it is difficult to picture the course of the resultant vector R during the day. But with the data of Table IV three-dimensional models of the figure swept out by R have been constructed for six stations. In addition to what has already been said about the plan, fig. 1, and east elevation, fig. 2, of these models, the remaining features of importance may be summarized as follows. At Seddin, De Bilt, and Eskdalemuir the vector figure is roughly circular and described almost wholly in a plane which, on the average of the whole year, is nearly vertical and normal to the magnetic meridian. As fig. 2 indicated, this plane is actually tilted slightly away from the vertical to the north at Seddin, almost exactly vertical at De Bilt, and towards the south at Eskdalemuir.

North of Eskdalemuir the form of the vector figure is fundamentally changed. The separate hourly vectors now tend to cluster about two directions represented approximately by the morning and evening maximum vectors. These two directions, one to northward and one to southward of the station, lie nearly in the plane of the local magnetic meridian, but they are not colinear either in azimuth or inclination to the vertical. The back of the figure is broken at the transition from north-directed to south-directed vectors. In addition, the general slope of the figure, while remaining nearly in the meridian, changes with latitude in the manner shown by fig. 2. The concentration of all hourly vectors about two directions is most pronounced at Sodankylä, for which station the figure is not more than 20° from the horizontal.

Farther north at Fort Rae these general features are maintained; but whereas in stations of zone z_2 the general slope of the figure is southward from the vertical, at Fort Rae the slope is towards the north. In addition, the separate vectors at Fort Rae tend to open out again fan-wise in the plane of tilt.

Fig. 4 represents diagrammatically the seven vector figures considered as if aligned along a magnetic meridian and regarded from the east. The broken lines represent the vertical at each station. The significance of the double position for the direction of current flow will be considered shortly; in true proportion the height of the current would be such that De Bilt is just on its horizon.

8—If the direction of flow in the current system were only reversed between a.m. and p.m., the latitude (or horizontal distribution of current concentration) and azimuth remaining unchanged, the directions of the a.m. and p.m. vector clusters would be colinear. An examination of the orientations of the vector shows convincingly that the p.m. position of the current is systematically different from its a.m. position, the change being representable as a northward displacement from a.m. to p.m. and that this displacement is accompanied by a veer in azimuth.



A, Seddin; B, De Bilt; C, Eskdalemuir; D, Sitka; E, Lerwick; F, Sodankylä; G, Fort Rae.

FIG. 4—Orientation of principal vectors in meridan plane. (Height above earth's surface of lines of current flow and change in height from a.m. to p.m. exaggerated.)

The data of Table IV do not lend themselves easily to demonstration of these changes. If the resultant vectors R at all stations were grouped about two dominant directions as at Sodankylä, two average R's representing all but the transition hours could be taken to replace the 24 separate values. But in general the inclination θ varies simultaneously with the azimuth ϕ, so that an average value of θ or ϕ has no meaning in defining R. Trial of several methods for circumventing this difficulty led to limiting the examination to not more than three hours grouped about each of the two greatest R values when θ and ϕ are most nearly stationary. For stations below the latitude of Sitka, figs. 1 and 2 have been used for confirmation of inferences.

The values of R, θ, and ϕ at the two times each day when R is greatest are given in Table VII. In this table section *b* is intended to illustrate that R, θ, and ϕ for the single hour of greatest R may be practically identical with the average values of these quantities for not more than 3 hours grouped about the two maxima (section *a*). The averages will, however, be used throughout. R p.m. will be used to denote the vector which is either a primary or secondary maximum in the evening hours; R a.m. will denote the morning maximum vector.

TABLE VII—MAGNITUDE AND ORIENTATION OF THE TWO MAXIMUM RESULTANT FIELD VECTORS OF THE REGULAR DAILY DISTURBANCE VARIATION

Observatory	*a*—Average of 2 or 3 hours grouped about the maximum				*b*—Single hourly values at the time of the maximum			
	Local time	R	θ	ϕ	Local time	R	θ	ϕ
	hr	γ	°	°	hr	γ	°	°
Fort Rae	5–8	155	57	188	6	157	53	191
	15–18	121	115	27	16	124	110	30
Sodankylä	23–2	129	106	164	1	134	106	164
	15–18	110	71	348	17	114	72	348
Lerwick	0–2	69	138	146	1	69	135	147
	15–18	72	44	338	17	75	45	338
Sitka	2–5	86	138	193	4	90	137	193
	18–21	71	36	25	20	72	36	26
Eskdalemuir	0–4	26	154	—	2	27	152	—
	17–20	33	18	—	18	34	15	—

As an example of the interpretation of the data of Table VII, R a.m. at Fort Rae is oriented 8° west of south and 57° from the (downward) vertical; R p.m. is 27° east of north and is 25° above the horizontal. Referring these vector positions to the disturbing current producing the field, the former inferences are confirmed that the main body of the disturbing current lies to the south of Fort Rae, both p.m. and a.m., and that its direction is reversed from being east-west a.m. to west-east p.m. In addition, it may now be inferred that that part of the current system responsible for the field changes at Fort Rae has moved northward (or, alternatively, increased in height) and veered between morning and evening and that, accompanying this displacement, the distribution of current has altered so as to produce a diminished resultant field in the evening hours.

A similar daily northward movement and veer in the current producing the vectors R a.m. and R p.m. at Sodankylä and Sitka is shown by Table VII. For Lerwick, though the inclination of R a.m. above the south horizon is the same (48°) as at Sitka, the corresponding angle for R p.m. below the north horizon is 44° against 36° at Sitka. This and the similarity in magnitude of R a.m. and R p.m. at Lerwick tend to confirm earlier deductions that in most respects the disturbance field at Lerwick is more appropriate to that of a station of lower magnetic latitude.

TABLE VIII

Observatory		ϕ °	Declination °	ψ °
Fort Rae	a.m.	8	37	24
	p.m.	27		
Sodankylä	a.m.	−16	1	−27
	p.m.	−12		
Sitka	a.m.	13	30	21
	p.m.	25		
Lerwick	a.m.	−24	−14	−24
	p.m.	22		
Eskdalemuir	17–18 hr (summer)	−37	−17	−20

Table VIII affords a comparison between the azimuth, ϕ, of R irrespective of sense, the azimuth of the local magnetic meridian D and ψ, the azimuth of each station relative to the magnetic axis pole, all azimuths considered positive when measured from north to east. The value −37° for ϕ for Eskdalemuir is derived from the computed azimuth of R at 17–18 hr in the average summer months when alone, fig. 1 shows that the horizontal vectors there tend to group round one direction.

At Fort Rae and at Sitka, the two stations with large easterly declination, the relative configurations of ϕ, D, and ψ are surprisingly similar, while at Sodankylä and Lerwick the arrangement, though differing from the other two, is only slightly less concordant. The main features are that at Fort Rae and Sitka the a.m. and p.m. positions of R lie immediately on either side of ψ, while D is much farther east. R p.m. at both stations is at most 4° from ψ. The same is true of Lerwick with the addition that here R a.m. coincides with ψ. At Sodankylä R a.m. is nearest ψ and R p.m. is equidistant from ψ and D. At both these stations D is also the most easterly of the four azimuths. The primary deduction to be made from the table is that the disturbance current for all stations lies more nearly at right angles to the great circle joining each station with the magnetic axis pole than to the direction of the local surface magnetic meridian.

9—On the assumption—by now recognized as inadequate—that a comparatively concentrated current flow is responsible for the vector orientations and magnitudes discussed in the foregoing paragraphs, attempts were made by scale drawings (in which the earth's curvature was accurately allowed for) to determine the approximate position of the current system. Though intersection of the lines drawn perpendicular to the vectors to represent the elevation of the current line above the horizon of each station was not expected, it was hoped they would converge on a limited area indicating a diffuse distribution. But the trials have made it clear that the disturbance currents must be widely distributed both laterally and vertically. Moreover, results to be presented in § 10 confirm conclusions reached by Goldie, *loc. cit.*, that the average position of the current system probably varies from month to month and also with the degree of disturbance, so that even if the complex form alone of the current system did not nullify graphical methods, the heterogeneous nature of the data used here would introduce uncertainties.

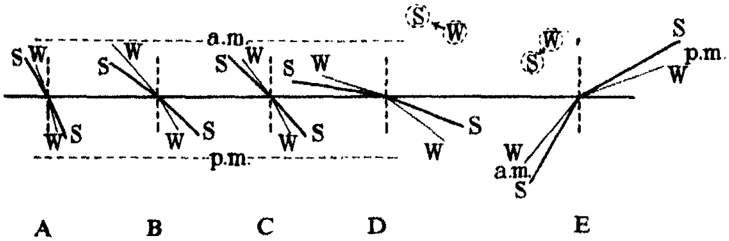
10—The behaviour of the disturbance vectors will now be examined for indications of seasonal changes in the current system. The appropriate details, extracted from unpublished tables for winter and summer similar to Table IV for the year, are given in Table IX. As far as possible the same hours as in Table VII have been selected for specification of the magnitude and orientation of R so long as those hours included the greatest value of R. An analysis of the table leads to the following conclusions:

(1) As diagrammatically represented in fig. 5, the direction of seasonal change of θ is the same at all stations and, on the hypothesis of a simple concentrated current over the latitudes between Sodankylä and Fort Rae, could be explained by a southward displacement of this current in summer. Even if the current were extensively diffuse, such a movement alone would also account for the increase from winter to summer in both the a.m. and p.m. field vectors at all stations south of the current line, except the a.m. Sitka vector, which is anomalously low. If, alternatively, the summer change of orientation and magnitude of R is to be ascribed to an increase

in height distribution of the current, this must be accompanied by an increased concentration of current flow. At Fort Rae to the north of the current line the alternative to a southward movement (which in turn would require to be accompanied by a substantial increase in current concentration) is a decreased mean height.

TABLE IX—A.M. AND P.M. RESULTANT DISTURBANCE VECTORS IN WINTER AND SUMMER

Observatory		Winter			Summer		
		R	θ	φ	R	θ	φ
		γ	°	°	γ	°	°
Fort Rae	a.m.	154	55	186	172	44	193
	p.m.	103	110	26	131	123	33
Sodankylä	a.m.	102	113	165	130	105	164
	p.m.	86	62	349	109	72	347
Lerwick	a.m.	49	143	142	72	140	145
	p.m.	53	36	338	73	49	334
Sitka	a.m.	87	145	189	68	137	200
	p.m.	59	31	32	73	41	17
Eskdalemuir	a.m.	18	155	—	32	152	—
	p.m.	27	10	—	41	29	—



A, Eskdalemuir; B, Sitka; C, Lerwick; D, Sodankylä; E, Fort Rae.

FIG. 5—Seasonal change of inclination of disturbance vector.

(2) The daily displacement or change in distribution of current already inferred from the data for the year as a whole in § 8 is confirmed. In both summer and winter the east-west, a.m., current is more southerly than the west-east, p.m., current. At Lerwick and Eskdalemuir this effect is noticeable only in winter.

As regards the rotation of the line of current flow, the data of Table IX show that at Fort Rae, Sodankylä, and Lerwick in both summer and winter the apparent northward displacement from the a.m. position is accompanied by a veer. At these three stations the amount of the rotation is slightly greater in winter than in summer. Lerwick, with a rotation in winter of 16° and in summer of 9°, shows the greatest seasonal change. Arising from the same anomaly as mentioned above, the veer of current at Sitka appears only in winter.

The difference in form of the three-dimensional vector figures for Eskdalemuir, De Bilt, and Seddin, compared with the clustered figure for the more northerly stations, makes it impracticable to bring the data for those stations into numerical alignment with the data of Table IX. But the separate component diagrams, especially those of fig. 2, have already shown, § 5, that the seasonal phenomena as regards southward movement, or increased height, in summer of the current system affecting those stations are equally clearly marked.

The daily mean values of R for winter and summer, Table X, show an unexpected grouping of stations in longitude rather than in latitude.

TABLE X

	Fort Rae	Sodankylä	Lerwick	Sitka	Eskdalemuir
	γ	γ	γ	γ	γ
Winter R	96	54	33	46	17
Summer R	97	72	44	47	24
Year R	99	70	43	52	21
Winter/summer	1·01	1·33	1·33	1·02	1·41

Whereas a similarity in the magnitude of the disturbance vector in the winter and summer months at Fort Rae may have been anticipated and is indeed confirmed by other criteria, the same feature at Sitka, which in most other disturbance characteristics groups naturally with Sodankylä and Lerwick, is unexpected. It provides a further indication that the disturbance field is asymmetrical with respect to the magnetic axis pole.

11—The effect on the magnitude and orientation of the disturbance vector on days of great as compared with average disturbance has been tested by the Fort Rae data alone. Using a local criterion, 40 of the most disturbed, d', and 38 of the quietest, q', days during the 13 months August 1932–August 1933 were segregated and difference $(d' - q')$ inequalities formed. From these the values of R, θ, and ϕ, given in Table XI, were computed.

Compared with the corresponding vector characteristics for $(d - q)$ days, Table IV, it is noticeable that, while the R p.m. maximum occurs at the same local hours, the greater R a.m. is advanced to 2–5 hr, though the three steadiest high values of R remain at 5–8 hr as for $(d - q)$ days. The daily mean R in Table X is 134 γ compared with 99 γ in Table IV.

TABLE XI—DISTURBANCE VECTOR MAGNITUDE AND ORIENTATION ON $(d' - q')$ DAYS AT FORT RAE

Local hour ending	R	θ	φ	Local hour ending	R	θ	φ
	γ	°	°		γ	°	°
1	150	57	240	13	116	100	30
2	122	40	232	14	132	102	33
3	170	42	217	15	158	110	34
4	168	28	200	16	172	113	31
5	170	29	193	17	162	121	29
6	162	42	192	18	159	127	27
7	163	40	181	19	146	146	22
8	163	48	181	20	151	162	28
9	98	62	180	21	133	169	355
10	39	96	168	22	130	165	251
11	28	125	92	23	99	138	251
12	81	108	30	24	148	96	241

Corresponding with Table VII for $(d - q)$ days, Table XII gives the mean values of R, θ, and φ for the two groups of 3 hours of greatest R; the means for the hours 2–5 hr, when the $(d' - q')$ vectors have their real maximum, are also shown. The $(d - q)$ data are included in the table for comparison. In addition to the increase of mean R, the effect of increased disturbance is to equalize R a.m. and R p.m. On $(d - q)$ days R a.m. exceeded R p.m. by 30% in winter and 49% in summer, Table IX.

TABLE XII

Fort Rae	$(d' - q')$			$(d - q)$		
	R	θ	φ	R	θ	φ
	γ	°	°	γ	°	°
a.m. (2–5 hr)	169	33	203	—	—	—
a.m. (5–8 hr)	163	43	185	155	57	188
p.m. (15–18 hr)	164	120	29	121	115	27

Increase of scale of disturbance at Fort Rae also affects the distribution of disturbance during the 24 hours of the day. In average disturbance the day and night mean resultant vectors are approximately equal, but in the group of most disturbed days R is 146 γ during the night and 123 γ during the day. The night/day ratio of ρ (0·60 for average disturbance) increases to 0·77 for the most disturbed days.

As regards inclination of the vector to the vertical, the change from $(d - q)$ to $(d' - q')$ days is similar to the change from winter to summer shown in the right-hand side of fig. 5. This implies that increased scale of disturbance at Fort Rae incurs southwards displacement of the current system (both a.m. and p.m.), which displacement must also be accompanied by an increased concentration of the lines of current flow, or the mean height must be lowered. On either hypothesis the change must be more marked for the east-west, a.m., current direction than for the p.m. current, west-east. This is necessary to account for the relative change both in ρ and in R.

The effect of increased disturbance on the azimuth of the current is not so unambiguous, though, when the values of ϕ are compared for the hours of real maxima in R (*i.e.*, 2–5 hr in $(d' - q')$ with 5–8 hr in $(d - q)$ and 15–18 hr for both) there is a veer from the less to the more disturbed days. As Table VIII shows, this implies that on the more highly disturbed days the current system in both its a.m. and p.m. positions is more nearly at right angles to the direction of local declination. If, as is likely, the increased field vectors on $(d' - q')$ days are the result of increased current as well as change in distribution of the lines of flow, this effect confirms another conclusion of Goldie, *loc. cit.*, that "the stronger the current, the more nearly does its direction approach to being perpendicular to the magnetic meridian."

SUMMARY

12—The three mutually perpendicular force vectors of the earth's magnetic field added during disturbance to those acting on quiet days are examined. From the characteristics in magnitude and orientation of these disturbing vectors at seven observatories it is found that the overhead current system producing the regular diurnal disturbing field in moderately high latitudes is primarily concentrated in a narrow zone asymmetrically encircling the magnetic axis pole and at about 23° from it. The direction of this current flow is east-west in the early local morning and a maximum in the reverse direction about 18 hr. A systematic daily change of space distribution and azimuth of the lines of current flow accompanies the change in current direction. In winter the disturbance vectors further indicate that the current zone lies more northerly than in summer, but the seasonal displacement, like the daily, is probably to be attributed more to a changed distribution in the vertical of the lines of current flow than to a horizontal movement of the whole system. At all times the disturbing system is diffuse and complex.

The Ultra-Violet Absorption Spectra of Some Complex Aromatic Hydrocarbons—I

By W. V. MAYNEORD, D.Sc., F.Inst.P., and E. M. F. ROE, B.Sc. (International Cancer Research Foundation Scholar), Physics Section of the Radiological Department, and Research Institute, The Cancer Hospital (Free), London.

(*Communicated by E. L. Kennaway, F.R.S.—Received May* 3, 1935)

INTRODUCTION

During earlier work carried out in this Institute on carcinogenic mixtures obtained from coal-tar, a study of the fluorescence spectra of such mixtures gave useful indications of the type of compound present,* and subsequently assisted in the separation of one of them, 1 : 2-benzpyrene. Very small quantities of material may often be detected by the study of fluorescence spectra, but the results are as a rule only qualitative. In general, at room temperature, the fluorescence spectra of hydrocarbons in solution consist of diffuse bands, and the total emitted energy is often very small and consequently difficult to measure. Moreover, the absolute amount of radiated energy depends on the energy distribution in the incident radiation, and on such experimental details as concentration of solutions, and size and shape of containing vessels.

On the other hand, a quantitative measurement of the ultra-violet absorption spectrum of a solution of a pure compound can be made relatively easily, and not only furnishes a method of identification of the substance but serves as a basis for a theoretical treatment of its molecular structure.

Recent progress in the synthesis of complex hydrocarbons has placed at our disposal a large number of these compounds in a pure state, and, in view of their considerable biological importance, it was thought desirable that a quantitative study of their optical properties should be made, both for purposes of identification and in the hope that the results might indicate some connexion between their molecular structure and biological action. The ultra-violet absorption spectra of a number of compounds structurally related to the carcinogenic hydrocarbons (themselves known to be related to the sterols, bile acids, and oestrogenic substances)* have therefore been examined. The results are here treated from a purely physical standpoint.

---

* Cook, Hieger, Kennaway, and Mayneord, 'Proc. Roy. Soc.,' B, vol. 111, p. 455 (1932).

GENERAL THEORETICAL RELATIONSHIPS

It is convenient to discuss the relationship between various quantities occurring in absorption spectroscopy. Consider a parallel beam of radiation incident perpendicularly on the surface of a solution to be investigated. Let I_0 be the intensity of the incident radiation. Then the intensity, I, transmitted through a thickness d of solution may be written,

$$I = I_0 \, 10^{-kd}; \tag{1}$$

k is then the "extinction coefficient" as usually defined, d being measured in centimetres. We may write $k = c\varepsilon$ where c is the concentration of the solution in gram-molecules per litre, and ε is generally known as the molecular extinction coefficient. We use the symbol ε throughout in this sense.

Evidently

$$\varepsilon = \frac{1}{cd} [\log_{10} I_0/I]. \tag{2}$$

The quantity in brackets is commonly called "Density" (D) and read directly from the spectrophotometer setting. In general, for the substances investigated, ε varies between such wide limits that $\log_{10} \varepsilon$ rather than ε is more conveniently plotted against wave-length.

The validity of the formula (2) is continually tested by the alternate use of different thicknesses of solution and different concentrations. No deviation outside the experimental error has been observed, and we conclude that the law holds for these very dilute solutions. From a purely physical point of view the notation implied in equation (1) is unfortunate, as absorption should clearly (as in other wave-length regions, for example, X-rays or γ-rays) be referred to base e (2·71828) rather than 10. The transfer is evidently made as below. Suppose ε' be the true molecular extinction coefficient as defined by the equation

$$I = I_0 \, e^{-c\varepsilon' d}, \tag{3}$$

where c and d have their previous significances. Then $\varepsilon' = 2{\cdot}303\,\varepsilon$ and $\log_{10} \varepsilon' = \log_{10} \varepsilon + 0{\cdot}3622$.

The molecular absorption coefficient μ_m (that is, referred to a single molecule) is then ε'/N, where N is the number of molecules per cc for a concentration of 1 gram-molecule in 1 litre ($N = 6 \cdot 06 \times 10^{20}$). μ_m rises for the molecules here considered to values of the order of 4×10^{-16}.

---

* Barry, Cook, Haslewood, Hewett, Hieger, and Kennaway, 'Proc. Roy. Soc.,' B, vol. 117, p. 318 (1935).

TECHNIQUE

All the substances have been investigated in solution at room temperature. Hexane is usually regarded as having least effect on the absorption spectrum of the solute, but as these complex hydrocarbons are relatively insoluble in this medium, ethyl alcohol has been used throughout, with the single exception of picene, which is extremely insoluble in alcohol and has therefore been examined in chloroform.

A standard Hilger E.316 all-metal quartz spectrograph has been used for the measurements. The dispersion of this instrument is such that the length of spectrum from 2000 to 4000 A is 177 mm. The source of light for the quantitative experiments has normally been a condensed tungsten steel spark, 4 mm gap, the necessary potential being obtained from an X-ray induction coil with mercury interrupter in the primary. The solutions are examined in a micrometer Baly tube, so that the optical path in the solution is variable continuously from 0·01 mm to 5·00 mm. Since a number of the substances are sparingly soluble even in alcohol, additional measurements have been made when necessary with 100 mm absorption cells.

Owing to the fact that with a line source narrow and feeble bands are easily completely overlooked, we have recently carried out qualitative examinations of all solutions, using as source of light a hydrogen discharge tube run at about 3000 volts derived from an A.C. transformer. On the continuous spectrum of this source, absorption detail (which is difficult to obtain with the quantitative set-up) may often be seen quite clearly.

Three forms of Hilger spectrophotometer have been used, namely (1) a rotating sector photometer, (2) the notched echelon cell outfit,* and (3) the "Spekker" photometer.†

A careful comparison of the accuracy obtained by the three methods was made for pyrene, chosen as a test substance on account of its representative and remarkable series of sharp absorption bands. The notched echelon cell outfit was found to be quite inadequate for the purpose in hand, since a high capacity for resolving detail of absorption bands was required. As the match points observed with this cell are at large wavelength intervals, small bands in regions of rapidly varying absorption coefficients are easily overlooked. To obtain accurate curves a great number of concentrations would have to be examined. Even if this difficulty were overcome, that of determining accurately a large number of match points on a small area of photographic plate would still be acute.

---

* Twyman, 'Proc. Phys. Soc.,' vol. 45, p. 1 (1933).

† Twyman, 'Trans. Opt. Soc.,' vol. 33, p. 1 (1931-32).

The rotating sector photometer was found to give much more reliable results, though it is still slightly less accurate than the "Spekker."* The latter is, moreover, considerably quicker and more convenient, and has therefore been used for all the experiments described subsequently.

The double spectra were recorded on Ilford Process plates and developed in a standard metol-hydroquinone developer, the plates being somewhat over-developed and subsequently reduced to obtain maximum contrast. In the far ultra-violet below about 2300 A we have used "Autofilter" plates, owing to their higher sensitivity. Where the 2000 A region is of particular interest these plates have been further sensitized in a solution of transformer oil in pentane. Possibly owing to the shorter optical path in quartz and the consequent smaller light absorption, the rotating sector photometer proves more suitable in the short wave-length region below 2300 A than the "Spekker" photometer, which also suffers from light losses during reflection. In general a wave-length scale was photographed twice on a given plate and the errors of the scale in different regions were estimated at the commencement of the experiments by comparison with the spark spectra of pure copper and iron. Since the errors were found to be, as a maximum, 2–3 A in the region of 3500 A this is in general of little account in view of the difficulty in estimating the position of bands to a higher accuracy.

The accuracy of the determination of the position of match-points depends greatly on the rate of change of absorption in the region considered. If $d\epsilon/d\lambda$ is small it is difficult to determine the position to an accuracy greater than 5 A. Fortunately, in the substances investigated $d\epsilon/d\lambda$ is often high ($d(\log \epsilon)/d\lambda$ of the order of $0\cdot25$ in benzene); the maxima of bands may be obtained to about 2 A in the short wave-length region, and 5 A at longer wave-lengths where the dispersion is lower.

Owing to the small quantities of materials available (occasionally only a few milligrammes) in a number of instances the relative values of ϵ for different wave-lengths are somewhat more accurate than the absolute values. The materials have been weighed on a micro-balance and the concentrations estimated to be accurate to about 2%. The final possible error in extinction coefficient is considered to be normally of the order of ± 5%.

PURITY OF MATERIALS

With few exceptions all of the compounds studied were prepared synthetically in this Institute by methods which are described in the original memoirs, to which reference is made. The specimens employed for spectroscopic purposes were the original highly purified analytical samples, or else specimens which had been specially prepared and purified for examination. So far as possible care was taken to avoid the use of polycyclic aromatic hydrocarbons of coal tar origin, as such substances are notoriously difficult to free completely from small amounts of contamination which may have an appreciable influence on the absorption spectra. The methods of preparation or other sources of material were as follows:

---

* Twyman, 'Trans. Opt. Soc., vol. 33, p. 1 (1931–32).

Benzene—Supplied by British Drug Houses, Ltd., "extra pure for molecular weight determination."

Naphthalene—Supplied by British Drug Houses, Ltd., "extra pure for molecular weight determination."

Anthracene—Dehydration of pure 9: 10-dihydro-anthraquinol, prepared by reduction of synthetic anthraquinone.

Phenanthrene—(*a*) Selenium dehydrogenation of 4-keto-1: 2: 3: 4-tetrahydrophenanthrene, followed by purification through the picrate.

(*b*) Platinum dehydrogenation of *as*. octahydrophenanthrene, Cook and Hewett (' J. Chem. Soc.,' p. 1098 (1933)).

1: 2-*cyclo*-Pentenophenanthrene—Cook and Hewett (' J. Chem. Soc.,' p. 1098 (1933)).

4: 5-Benzhydrindene—Cook and Hewett (' J. Chem. Soc.,' p. 1098 (1933)).

Saturated dimeride of 1-*iso*Propenylnaphthalene—Cook (' J. Chem. Soc.,' p. 456 (1932)).

Chrysene—Cook and Hewett (' J. Chem. Soc.,' p. 372 (1934)).

Picene—Ruzicka and Hösli (' Helv. Chim. Acta,' vol. 17, p. 470 (1934)).

Pyrene—The " pure pyrene " supplied by the Gesellschaft für Teerverwertung was freed from its coloured impurity by partial hydrogenation, and then purified by repeated recrystallization of the hydrocarbon and its picrate.

2-Methylpyrene—The " 1-methylpyrene " of Cook and Hewett (' J. Chem. Soc.,' p. 365 (1934)). The true position of the methyl group was shown by direct comparison with a specimen kindly supplied by Dr. G. Kränzlein, who synthesized it by a method which establishes its constitution.

1: 2-Benzanthracene—The yellow material prepared by the method of Barnett and Matthews (' Chem. News,' vol. 130, p. 339 (1925)) was freed from colouring matter by agitation of a benzene solution with sulphuric acid followed by repeated crystallization of the picrate. The pure hydrocarbon formed colourless leaflets, m.p. 160° C, with a violet fluorescence.

6-Methyl-1 : 2-benzanthracene—Cook (' J. Chem. Soc.,' p. 456 (1932)).

7-Methyl-1 : 2-benzanthracene—Cook (' J. Chem. Soc.,' p. 456 (1932)).

6 : 7-Dimethyl-1 : 2-benzanthracene—Cook (' J. Chem. Soc.,' p. 456 (1932)).

10-*iso*Propyl-1 : 2-benzanthracene—Cook (' J. Chem. Soc.,' p. 456 (1932)).

5 : 6-*cyclo*Penteno-1 : 2-benzanthracene—Cook (' J. Chem. Soc.,' p. 2529 (1931)).

6 : 7-*cyclo*Penteno-1 : 2-benzanthracene—Cook (' J. Chem. Soc.,' p. 2529 (1931)).

1 : 2 : 5 : 6-Dibenzanthracene—Clar (' Ber. deuts. chem. Ges.,' vol. 62, p. 350 (1929)) purified, as described by Cook, Hieger, Kennaway, and Mayneord (' Proc. Roy. Soc.,' B, vol. 111, p. 455 (1932)).

1 : 2 : 5 : 6-Dibenz-9 : 10-anthraquinone—Cook (' J. Chem. Soc.,' p. 1477 (1932)).

9 : 10-Dihydro-1 : 2 : 5 : 6-dibenzanthracene—Cook (' J. Chem. Soc.,' p. 1592 (1933)).

9 : 10-Dimethyl-1 : 2 : 5 : 6-dibenzanthracene—Cook (' J. Chem. Soc.,' p. 489 (1931)).

cis- and *trans-* 9 : 10-Dimethyl-9 : 10-dihydro-1 : 2 : 5 : 6-dibenzanthracenes—Cook (' J. Chem. Soc.,' p. 489 (1931)).

1 : 2-Benzpyrene—Cook and Hewett (' J. Chem. Soc.,' p. 398 (1933)).

Methylcholanthrene—Cook and Haslewood (' J. Chem. Soc.,' p. 428 (1934)).

2′ : 1′-Naphtha-1 : 2-fluorene—Cook, Hewett, Mayneord, and Roe (' J. Chem. Soc.,' p. 1727 (1934)).

5-Methyl-8-*iso*propyl-2′ : 1′-naphtha-1 : 2-fluorene—Cook, Hewett, Mayneord, and Roe (' J. Chem. Soc.,' p. 1727 (1934)).

Neoergostatriene—Bonstedt (' Z. physiol. Chem.,' vol. 185, p. 165 (1929)).

Neoergostatetraene—Haslewood and Roe (' J. Chem. Soc.,' p. 465 (1935)).

Neoergostapentaene—Haslewood and Roe (' J. Chem. Soc.,' p. 465 (1935)).

EXPERIMENTAL RESULTS

Many hundreds of papers have appeared on the absorption spectra of organic compounds, but much of the early work was qualitative, whilst many of the compounds quoted here have only recently been synthesized, and their optical properties have therefore not previously been examined. Results relevant to a few of the compounds have been published by other workers, and our own results for these substances are included for completeness.

Figs. 1 to 19 show the variations of $\log_{10} \varepsilon$ with wave-length for the series of hydrocarbons investigated. We may remark that all points plotted are experimentally determined values. For convenience in description the curves have been divided somewhat arbitrarily into groups arranged according to the parent hydrocarbon.

Benzene Group

As might be expected from its fundamental importance, the absorption of benzene has been studied by a number of observers, notable among whom is Henri.* Our results, fig. 1, are in good agreement with his values, due allowance being made for the fact that Henri's results were obtained in hexane solution and our own in ethyl alcohol. Small differences appear in the region of 2380 A where a " shoulder " in the Henri curve has been further resolved. An additional band at 2240 A has been observed. On the long wave-length side of five of the main absorption bands " shoulders " have been observed by both of us at room temperature. We have recently commenced the study of absorption spectra of solutions at low temperatures, and at −183° C these "shoulders" in the benzene spectrum are easily resolved into separate bands, as with solid benzene† at low temperatures.

It is interesting to observe that three other substances examined, namely, neoergosta-triene, -tetraene and -pentaene, fig. 2, all much heavier and more complex molecules, show absorption spectra similar to benzene, but with suppressed detail.‡

Naphthalene Group

The curve for naphthalene is shown in fig. 3 and may be compared with the results of de Làszlò,§ and others,‖ who examined the spectrum in hexane solution. The spectrum may be divided into three absorption systems and extends farther into the visible than is the case with benzene. For an attempted detailed analysis of the vapour absorption spectrum and its relation to the spectrum in solution, we may refer to the work of Henri and de Làszlò.* The analysis evidently cannot be regarded as final in view of the serious discrepancy between the moment of inertia of the naphthalene molecule, as determined spectroscopically, and the value deduced from X-ray diffraction methods. Similar difficulties arise for benzene, and in both cases the spectroscopic data lead to much smaller values of the moment of inertia.

---

* 'J. Phys. Rad.,' vol. 3, p. 180 (1922).

† Kronenberger and Pringsheim, 'Z. Physik,' vol. 40, p. 75 (1926).

‡ Haslewood and Roe, 'J. Chem. Soc.,' p. 465 (1935).

§ 'Z. phys. Chem.,' vol. 118, p. 369 (1925).

‖ Morton and de Gouveia, 'J. Chem. Soc.,' p. 916 (1934).

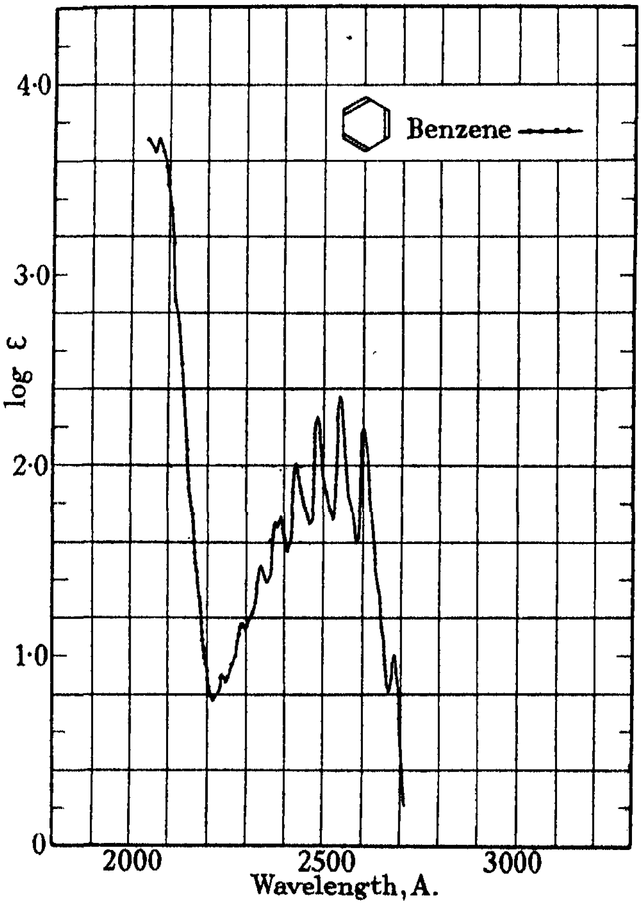


Fig. 1.

The addition of a five-membered ring (*cyclo*penteno), as in 4:5-benzhydrindene, fig. 4, has only a slight effect on the position of the band systems, but the prominence of the longer wave group is enhanced. The other two band systems remain substantially unaltered. In the polymerized *iso*propenyl-naphthalene detail is further suppressed, fig. 4. It is interesting to record the similarity of the absorption spectra of *cis*- and *trans*- 9: 10-dimethyl-9: 10-dihydro-1: 2: 5: 6-dibenzanthracenes, fig. 5, mentioned again later, to that of benzhydrindene. The saturation of the 9: 10-positions in the central dibenzanthracene ring appears to give rise

---

* 'Proc. Roy. Soc.,' A, vol. 105, p. 662 (1924).

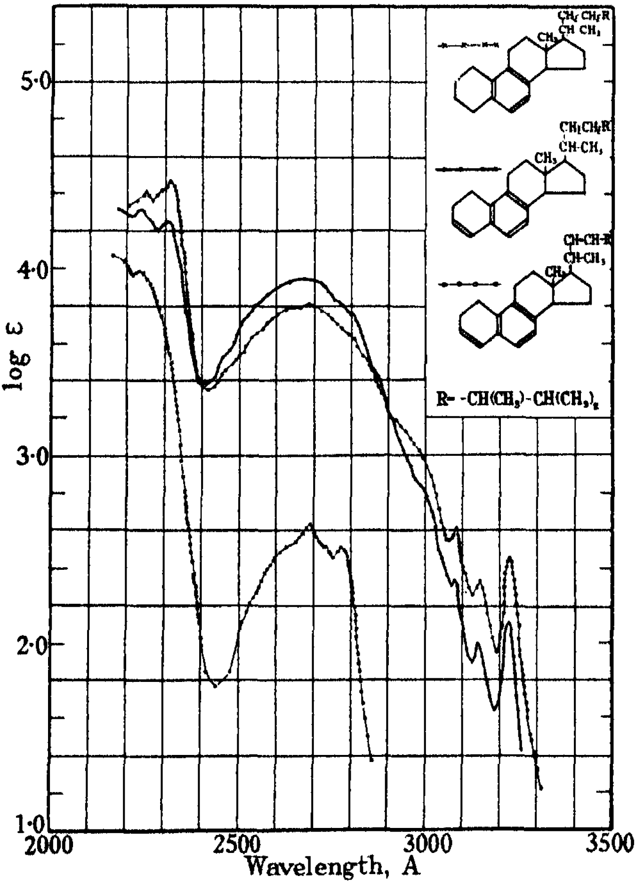


FIG. 2—× Neoergostatriene; ● neoergostatetraene; O neoergostapentaene.

to systems similar to naphthalene or benzhydrindene at either end of the molecule.

Anthracene Group

The absorption spectrum of anthracene is shown in fig. 6. A large shift to longer wave-lengths, as well as a greatly increased band-width compared with benzene, is apparent. The results are in accord with those of Clar,* but the extinction coefficients found by Radulescu† are lower, and of Capper and Marsh‡ (chloroform solutions) higher than here reported. The increased band-width as compared to benzene or naphthalene is noteworthy. As the " shoulders " now appear on the short wave-length side of the main bands, whereas for benzene they appear on the long wave-length side, it is evident that the relative intensities of the two components of each complex band have been reversed. At low temperatures preliminary observations have again shown separation of these components.

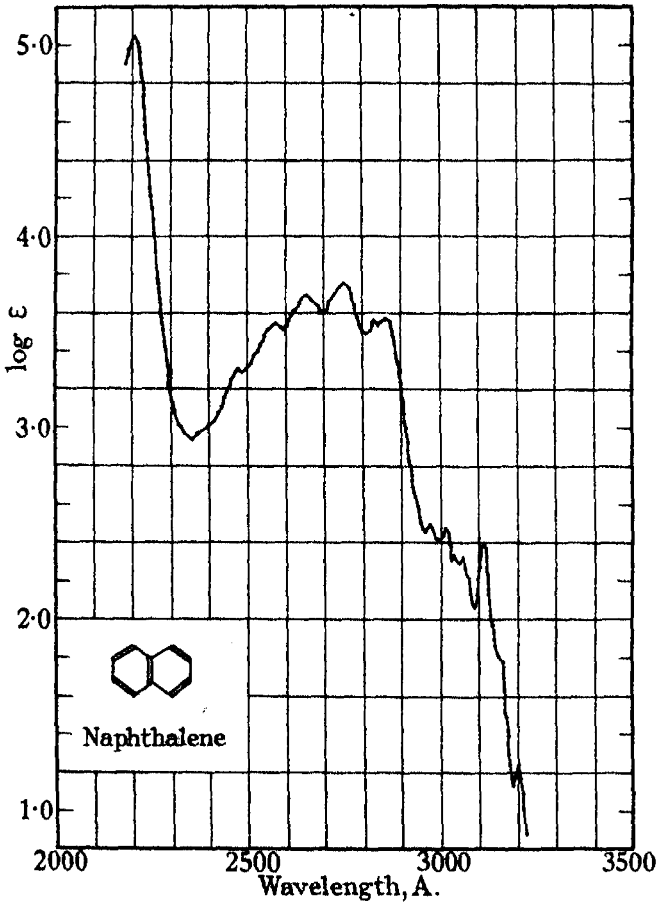


FIG. 3.

---

* ' Ber. deuts. chem. Ges.,' B, vol. 65, p. 503 (1932).

† ' Ber. deuts. chem. Ges.,' B, vol. 64, p. 2223 (1931).

‡ ' J. Chem. Soc.,' p. 724 (1926).

The striking effect of an additional ring in the 1 : 2-position is shown in the spectrum of 1 : 2-benzanthracene, fig. 7, previously examined in less detail by Clar (*loc. cit.*), and Capper and Marsh (*loc. cit.*), while the addition of a further methyl group has very little effect, as, for example, in 7-methyl-1 : 2-benzanthracene, fig. 8. It is interesting to observe that the groups of bands centred at 3400 A and 2800 A respectively are lowered by the introduction of a methyl group in the 7-position, while the band at 3840 A in benzanthracene is raised. By contrast, fig. 8, in 6-methyl-1 : 2-benzanthracene the latter band is considerably depressed, while the main band systems are again depressed, as in the 7-methyl compound. In the latter the band at 2994 A (in 1 : 2-benzanthracene) has approximately the same extinction coefficient, whereas it is raised in the 6-methyl analogue. The

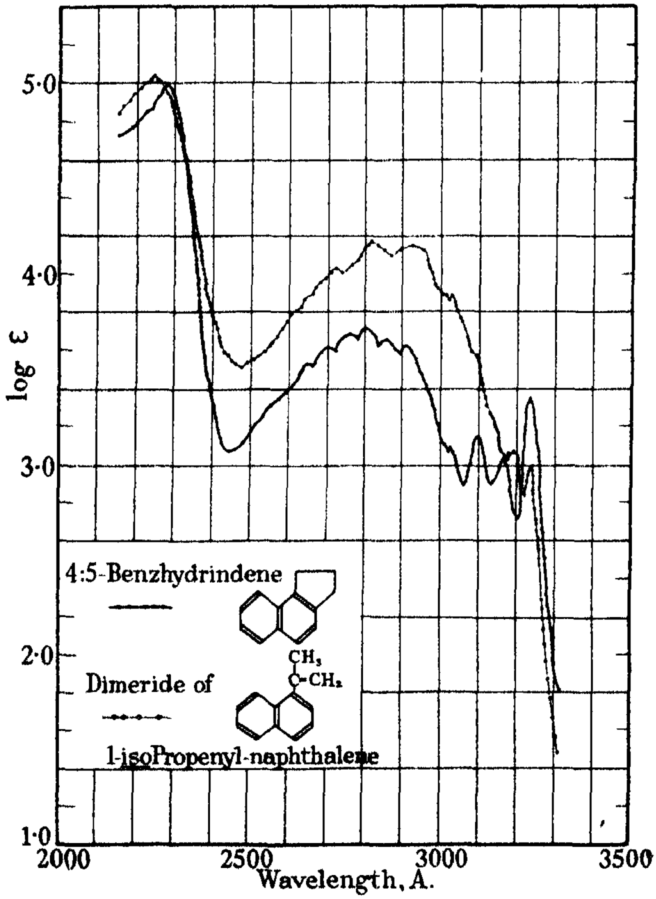


FIG. 4.

simultaneous inclusion of methyl groups in the 6- and 7-positions (6 : 7-dimethyl-1 : 2-benzanthracene, fig. 9) does not give merely additive changes in extinction coefficients but results in a spectrum very similar to that of 6 : 7-*cyclo*penteno-1 : 2-benzanthracene, fig. 9. There are, however, slight differences; for example, the bands 3029 A and 3855 A in the dimethyl compound are less pronounced than in the cyclopenteno compound, whereas the reverse occurs for the band at 2690 A. Rather less detailed results, published by Rosenheim and King,* show somewhat different relative intensities in the band systems observed for these two compounds.

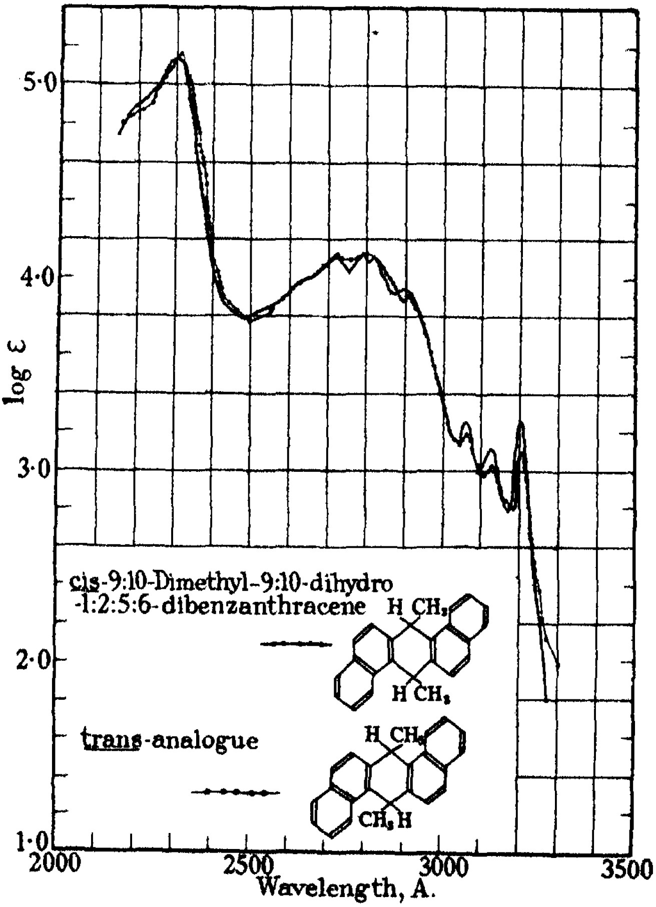


FIG. 5.

In the 5 : 6-*cyclo*penteno-1 : 2-benzanthracene, fig. 10, marked differences appear compared with the corresponding 6 : 7-*cyclo*penteno-1 : 2-benzanthracene. There is a general movement of the band systems towards the visible in the 5 : 6-*cyclo*penteno compound, while there are marked differences in the behaviour of bands at 3030 A and 3855 A, the bands being considerably more pronounced in the 6 : 7-compound. These two small bands which appear on the long wave-length slopes of the main systems in the benzanthracene compounds appear to behave independently of those systems. They may be analogous to the corresponding absorption band at 2682 A in benzene, whose frequency difference from its neighbours is known to be anomalous, and whose behaviour at low temperatures we have also recently observed to be different from that of the rest of the bands.

---

* 'J. Soc. Chem. and Ind.,' vol. 52, p. 299 (1933).

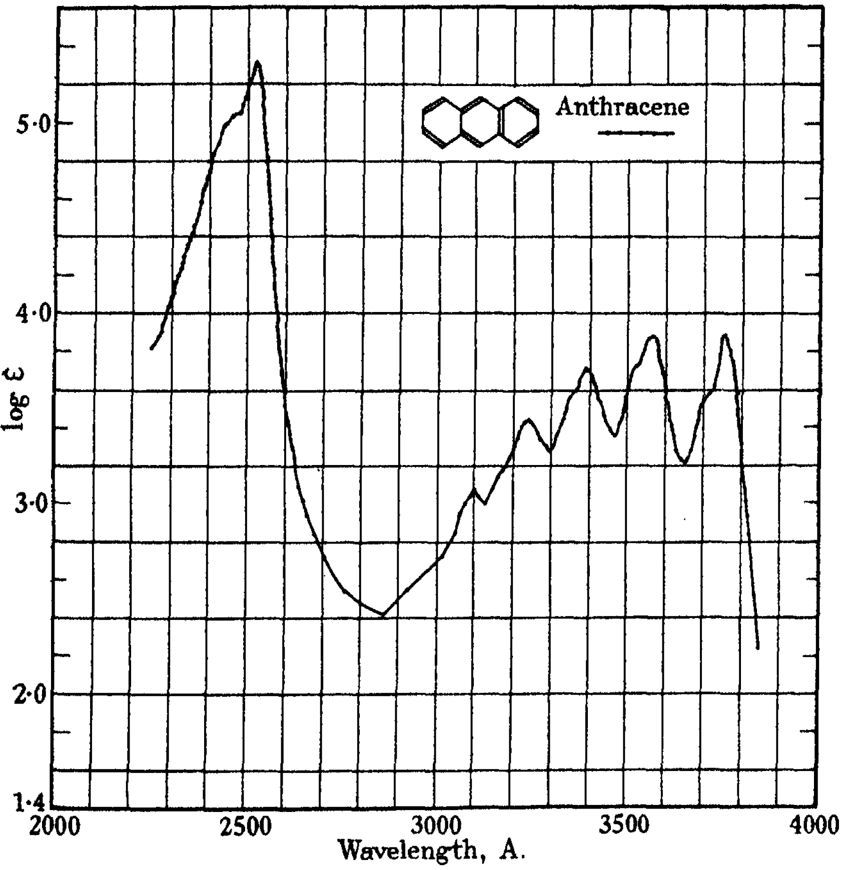


FIG. 6.

Two other derivatives of 1 : 2-benzanthracene have been investigated and show very similar absorption spectra, fig. 11. This is the more interesting in view of the fact that one of them, methylcholanthrene, is strongly carcinogenic while the other, 10-*iso*propyl-1 : 2-benzanthracene, is quite inactive.

These spectra again show the distinctive behaviour of the small bands discussed above. They are more intense in methylcholanthrene than in 10-*iso*propyl-1 : 2-benzanthracene, although all other bands show the reverse order of intensity.

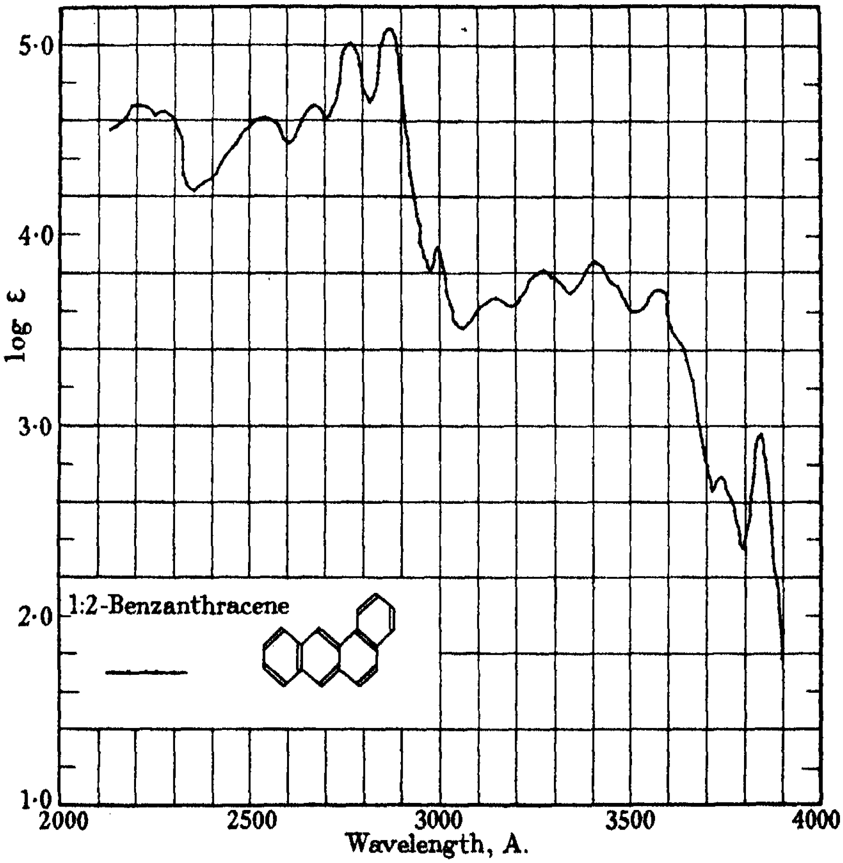


FIG. 7.

Phenanthrene Group

The absorption of phenanthrene is shown in fig. 12. Compared with the results of other observers,*†‡ we have found two new bands at 2195 and 2100 A respectively. It is unlikely that these are due to impurities as similar new bands in the far ultra-violet have been observed for a number of other compounds.

---

* Radulescu, 'Ber. deuts. chem. Ges.,' B, vol. 64, p. 2223 (1931).

† Capper and Marsh, 'J. Chem. Soc.,' p. 724 (1926).

‡ Clar and Lombardi, 'Ber. deuts. chem. Ges.,' B, vol. 65, p. 1411 (1932).

The beautifully regular long-wave system of bands is a prominent feature of the phenanthrene spectrum. We are reminded of the similar system in 4: 5-benzhydrindene where substitution in the 1: 2-position of the naphthalene molecule enhanced the intensity of the corresponding components of the naphthalene absorption spectrum. In chrysene, fig. 14, which may also be regarded as a 1: 2-substituted naphthalene, the long-wave system is again similar. Frequency differences of the order of 700 cm^{-1} are observed in the absorption spectra of phenanthrene, benzhydrindene, and chrysene for this group of bands. The frequency differences between neighbouring bands in phenanthrene, in fact, alternate

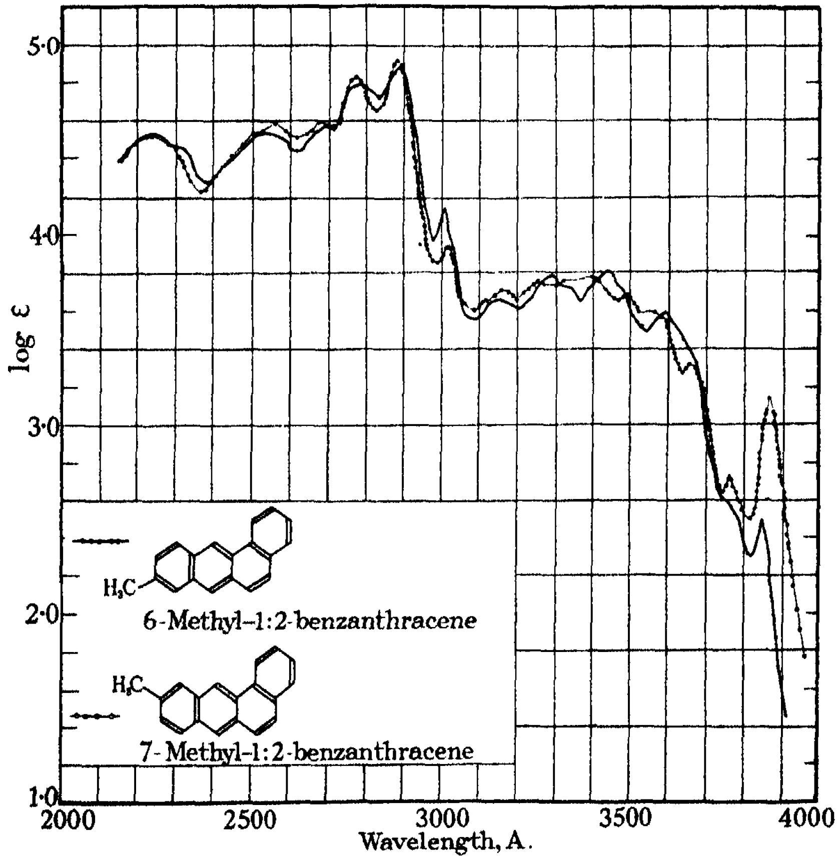


FIG. 8.

in value between two limits of approximately 650 and 750 cm^{-1}, suggesting that the system is double. Further evidence in support of this suggestion is obtained from the absorption spectrum of 1 : 2*cyclo*-pentenophenanthrene, fig. 12, where the addition of a five-membered ring in the phenanthrene 1 : 2-position emphasizes or depresses alternate bands.

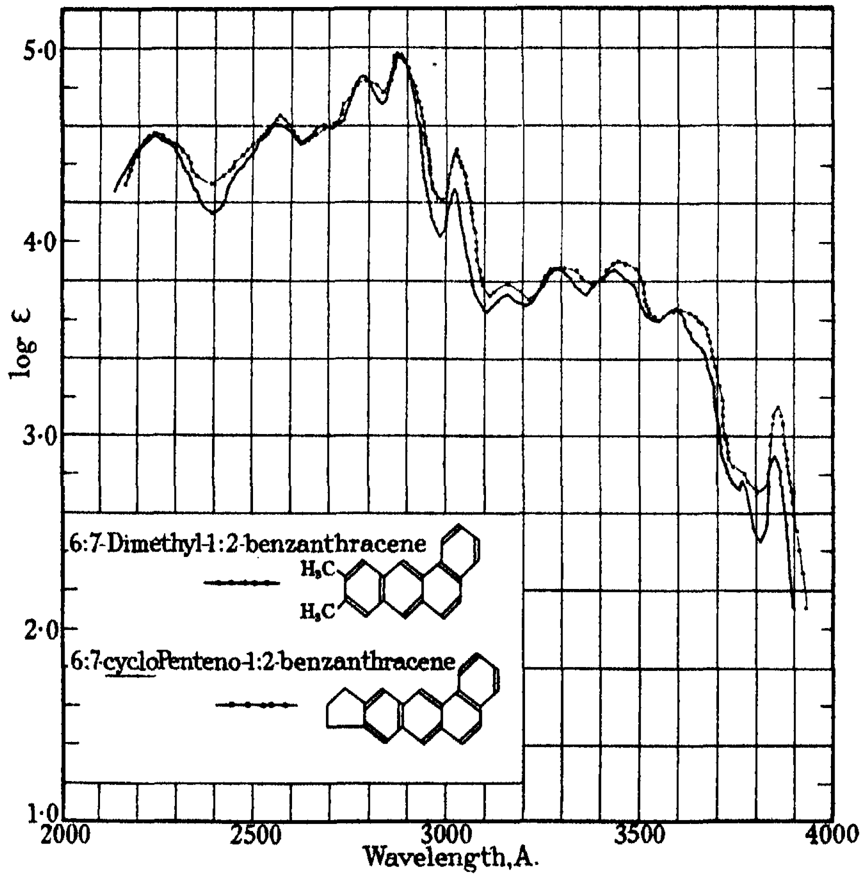


FIG. 9.

A further benzene ring added to 1 : 2-*cyclo*penteno-phenanthrene, as in 2′ : 1′-naphtha-1 : 2-fluorene, fig. 13, suppresses detail in the short-wave region, while the shift to longer wave-lengths has apparently engulfed two further bands of the long-wave system of phenanthrene. The latter are now just visible as " shoulders " in the region of 3300 A. In the 5-methyl-8-*iso*propyl derivative of naphthafluorene (shown for comparison in fig. 13) slight suppression of detail and shift to longer wave-lengths is observed.

The absorption spectrum of chrysene, fig. 14, also examined in less detail by Clar,* may be compared with that of 1 : 2-*cyclo*pentenophenanthrene. The long-wave system still shows its alternations of intensity, but movement towards longer wave-lengths of the central band-group has again obliterated some of the long wave bands.

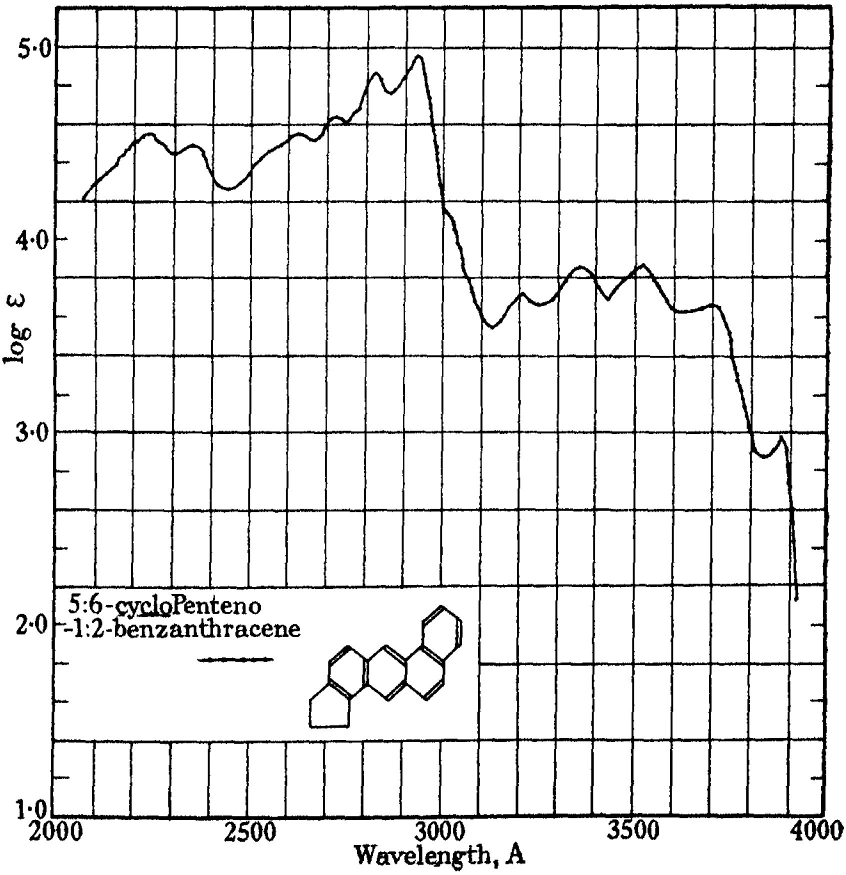


FIG. 10.

The absorption spectrum of picene, fig. 14, shows resemblances to the spectra of both phenanthrene and chrysene, of which the ring systems are included in the picene molecule. The long-wave band system strongly resembles that of chrysene. The central band system shows resemblances to phenanthrene and naphthafluorene, while the far ultra-violet bands of picene resemble the corresponding phenanthrene bands. Moreover, they include a third band, at 2575 A in picene, an analogue of which is observed in chrysene.

---

* Clar and Lombardi, 'Ber. deuts. chem. Ges.,' B, vol. 65, p. 1411 (1932).

Dibenzanthracene Group

The absorption spectrum of 1:2:5:6-dibenzanthracene, fig. 15, is surprisingly similar to that of chrysene, although the whole spectrum is

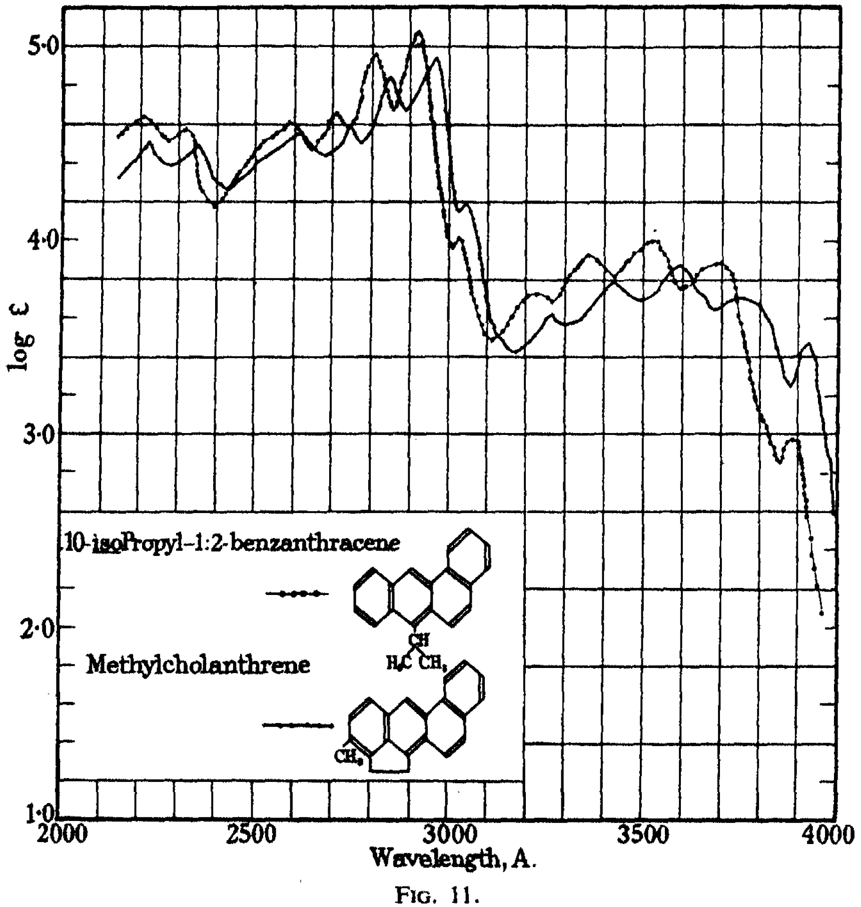


FIG. 11.

shifted to the limits of the visible in the former compound. The substitution of two methyl groups in the 9:10-positions considerably modifies the structure of the long wave-length band system, which is almost entirely obliterated (9:10-dimethyl-1:2:5:6-dibenzanthracene, fig. 15). There is a further considerable wave-length shift towards the visible, while a new band appears in the far ultra-violet at 2406 A; but the spectrum remains fundamentally similar to that of 1:2:5:6-dibenzanthracene.

On the other hand, in the 9 : 10-dihydro derivative, fig. 16, where the central dibenzanthracene ring is fully saturated in the 9 : 10-positions, the absorption is quite unlike that of any pure compound of which we are aware. A number of bands remains at approximately the same wavelength as in 1 : 2 : 5 : 6-dibenzanthracene, but their relative intensities are completely altered, resulting in a new general form of the curve. The absorption spectra of *cis-* and *trans-* 9 : 10-dimethyl-9 : 10-dihydro-1 : 2 : 5 : 6-dibenzanthracenes, fig. 5, reverted strikingly to the naphthalene and benzhydrindene type of curve. Here again the 9 : 10-positions are fully saturated, evidently dividing the molecule into two symmetrical portions each of which is naphthalenic. It is curious that the 9 : 10-dihydro-compound does not yield analogous results, although the purity of the specimen has been carefully investigated. The same curve was

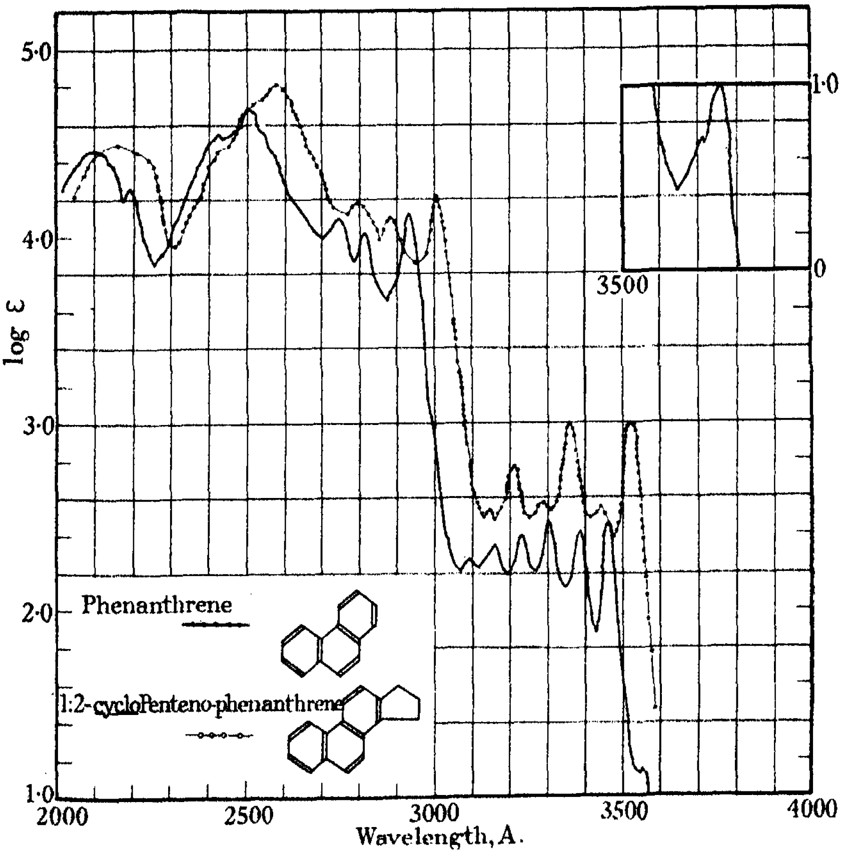


Fig. 12.

given by a specimen purified through the *s*-trinitrobenzene complex. We may mention that the absorption spectrum of 9 : 10-dihydroanthracene* is somewhat similar to that of benzene, confirming the general view that the molecule behaves as though consisting of two fairly separate portions when the central ring is appropriately saturated. The 1 : 2 : 5 : 6-dibenz-9 : 10-anthraquinone shows little fine structure, fig. 17, but a few broad bands extending into the visible as in benzoquinone.†

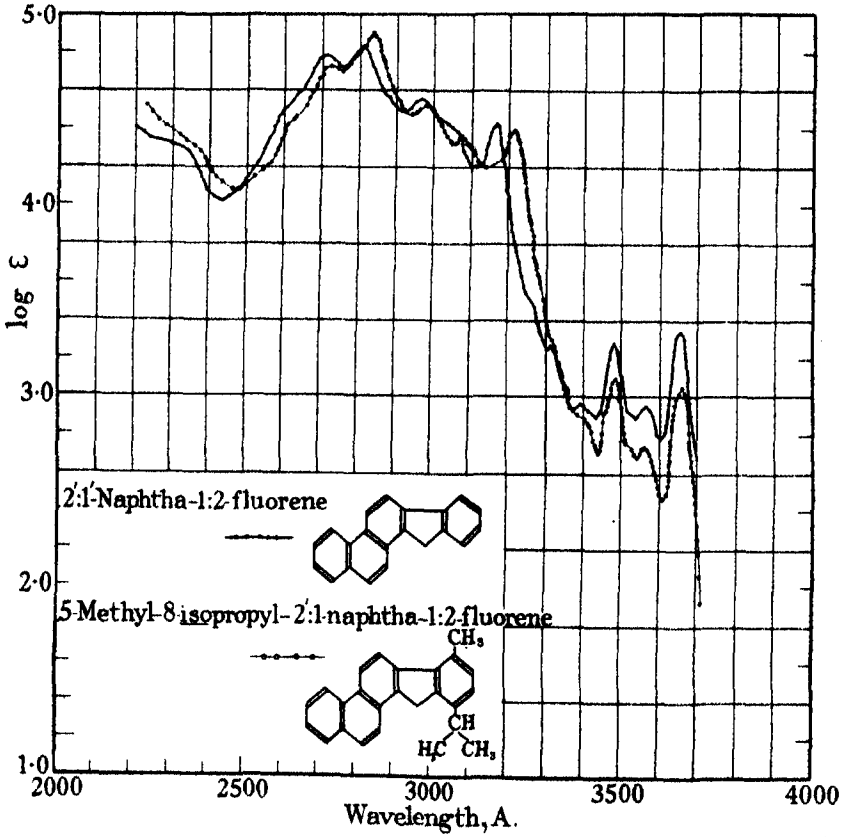


FIG. 13.

Pyrene Group

One of the most striking absorption spectra examined is that of pyrene,‡ fig. 18. It consists of a number of sharply defined maxima rising to high values of extinction coefficient. These bands appear to be divided into definite groups rather differently from those of any molecule previously investigated. The addition of a methyl group in the 2-position, 2-methyl pyrene, fig. 18, depresses the absolute value of the extinction coefficients and causes a slight shift to the visible, although every band is reproduced. This result brings strikingly to view the importance of the structure of the molecule as a whole in determining the form of the absorption spectra of these complex hydrocarbons.

---

* Barnett, Cook, and Ellison, 'J. Chem. Soc.,' p. 885 (1928).

† Klingstedt, 'C.R. Acad. Sci., Paris,' vol. 176, p. 1550 (1923).

‡ Clar, 'Ber. deuts. chem. Ges.,' B, vol. 65, p. 1425 (1932).

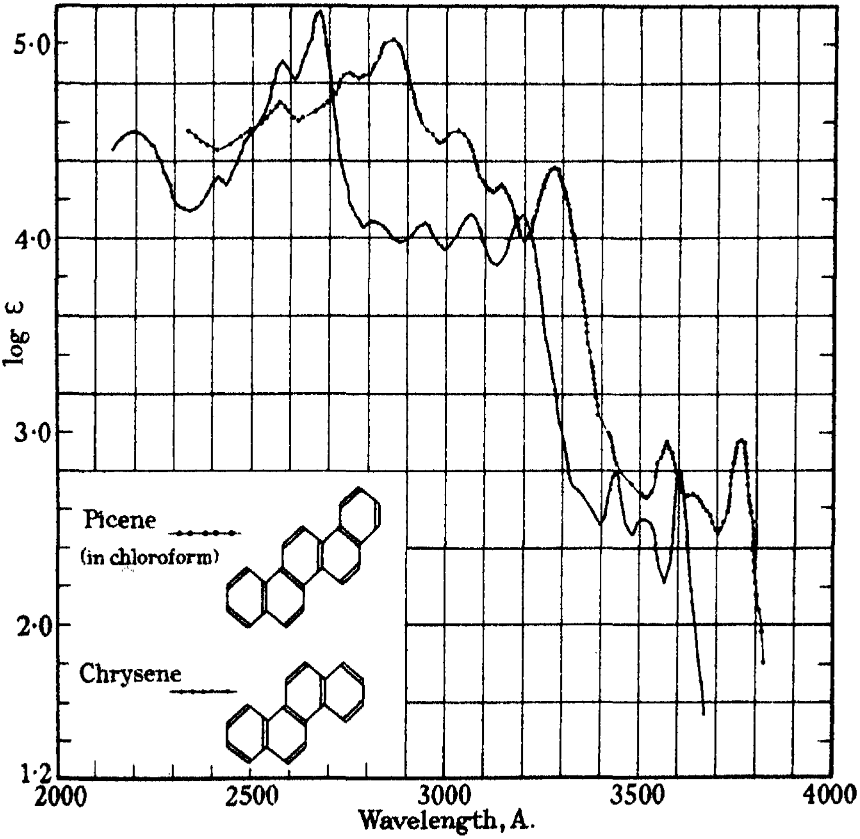


FIG. 14.

The addition of a further benzene ring in 1 : 2-benzpyrene, fig. 19, causes wave-length shifts, particularly in the group at about 3200 A in pyrene, the movement being only half as great in the short-wave system of bands. In the latter the shape of the curve is altered by the fusion of two band groups into one, so that a slight similarity to the spectrum of

1 : 2-benzanthracene, fig. 7, is now evinced. It is evident from the formula that the 1 : 2-benzpyrene molecule does contain the 1 : 2-benzanthracene nucleus.

REMARKS ON ABSORPTION SPECTRA

In view of the complexity of the molecules involved and the difficulties of the theoretical interpretation even for the simplest molecule, benzene, it is not surprising that any attempt to give a detailed explanation of these absorption spectra has proved unsuccessful.

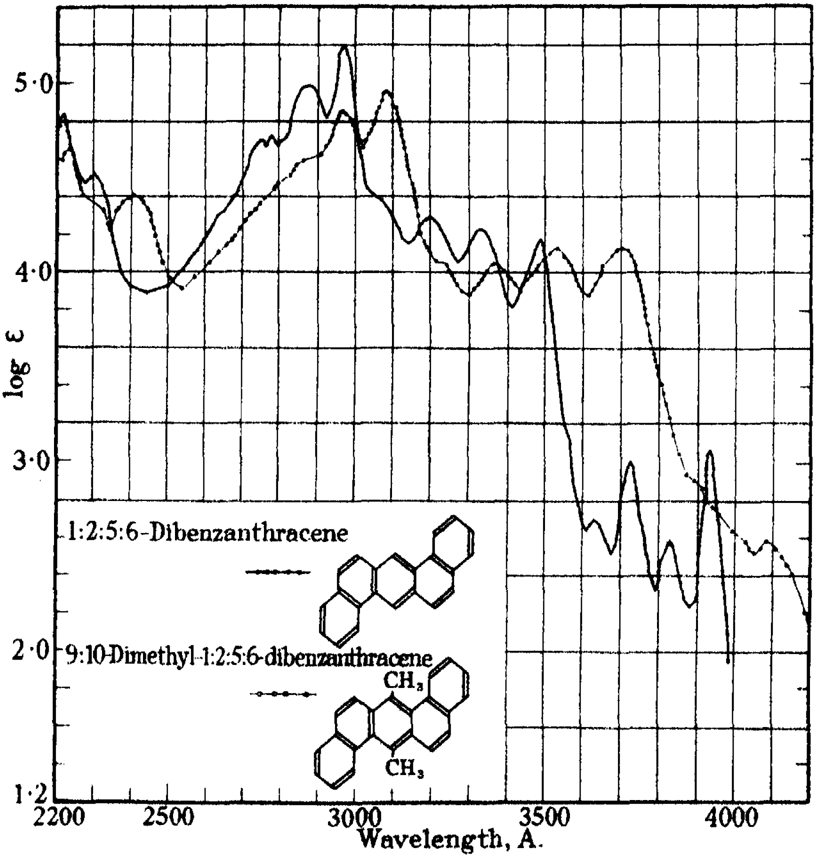


FIG. 15.

The spectra of benzene and naphthalene have been previously examined in detail by Henri and de Làszlò,[*] both in the vapour and dissolved state, and the general correspondence between the absorption in the two states

---

[*] 'Proc. Roy. Soc.,' A, vol. 105, p. 662 (1924).

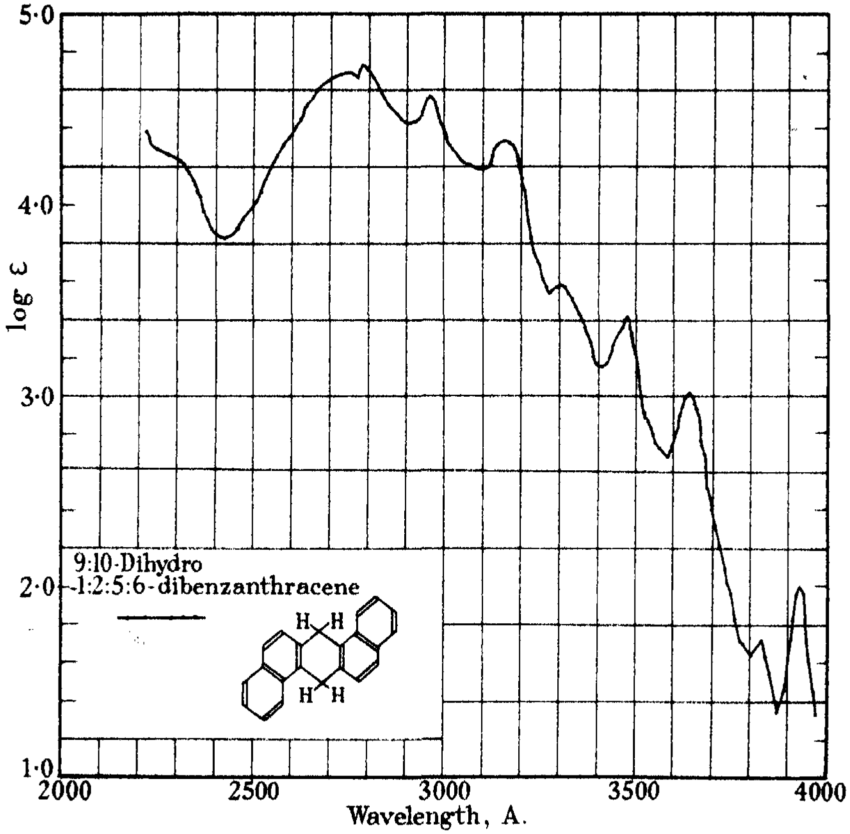

5·0
4·0
3·0
2·0
1·0
log ε
2000
2500
3000
3500
4000
Wavelength, A.
9:10-Dihydro
-1:2:5:6-dibenzanthracene
H
H
H
H

FIG. 16.

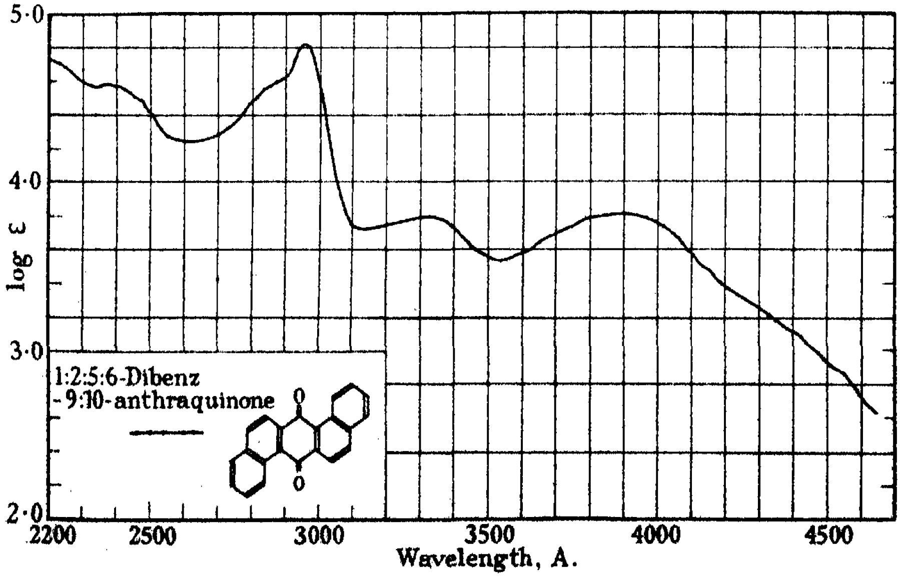

5·0
4·0
3·0
2·0
log ε
2200
2500
3000
3500
4000
4500
Wavelength, A.
1:2:5:6-Dibenz
-9:10-anthraquinone
O
O

FIG. 17.

made clear. The absorption spectra in solution show many of the characteristics of those of the vapour state; but fine detail, doubtless mainly due to rotational structure, is obliterated, the long-wave absorption bands in solution representing the envelopes of the line absorptions in the vapour state. In addition, wave-length shifts occur relative to the absorption in the vapour, in magnitude dependent on the solvent but of the order of 20 A. It is usually supposed that the different band systems correspond to different electronic excitation levels, the individual bands in the same system to varying atomic vibrational levels, while, as already mentioned, the rotational quantification is either hidden or distorted in the general outline of each band. It is, however, doubtful if for complex molecules such simple separation into three energy systems is valid. It may be noted that the wave-mechanical treatment of aromatic molecules has not reached a state which enables predictions of absorption spectra

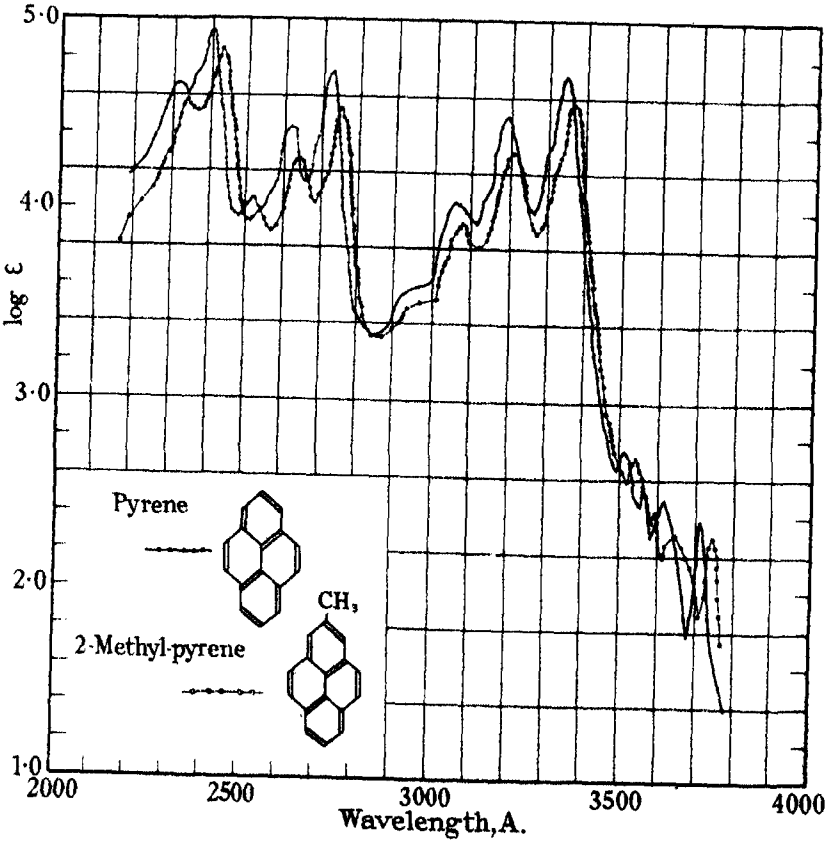


Fig. 18.

to be made, though interesting attacks on the general problem of molecular structure have been attempted, for instance, by Hückel,* in terms of the Hund π and σ electrons† involved in a double bond.

We have investigated the frequency difference between the individual band maxima of the substances examined, and for certain substances these differences may be correlated with frequency differences observed

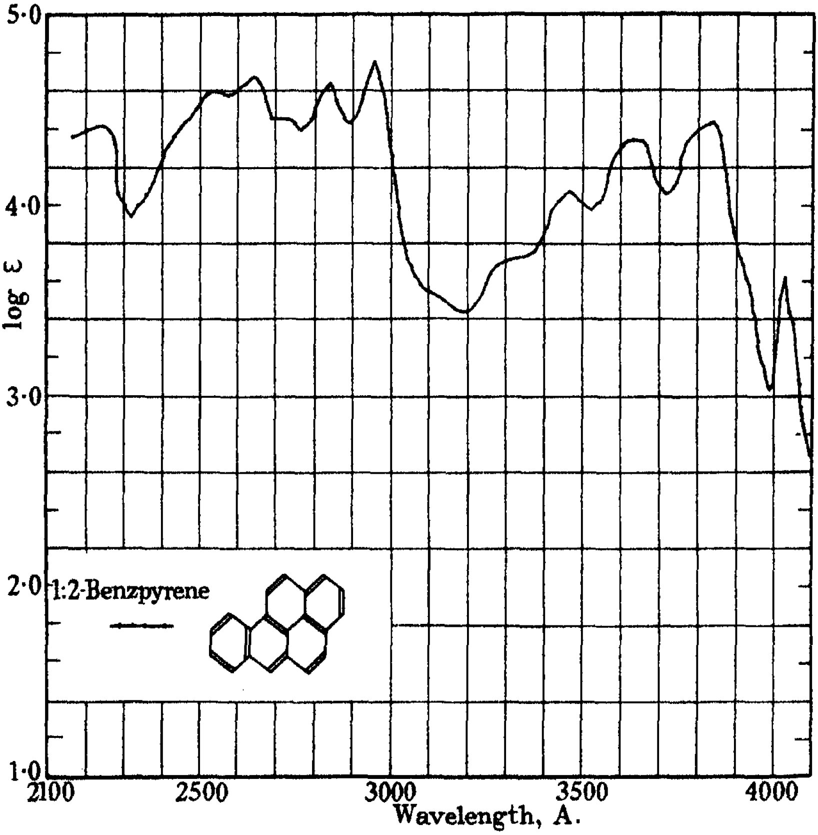


Fig. 19.

in Raman spectra. These again may be compared with frequencies found in observations on mechanical models of rings of atoms.‡

It is not proposed to discuss this further at the moment, as low-temperature investigations in progress seem likely to yield more suitable material for analysis.

---

* ' Int. Conf. Phys.,' vol. 2, p. 9 (1935).

† Hund, ' Z. Physik,' vol. 73, pp. 1, 565 (1931-32).

‡ Kettering, Shutts, and Andrews, ' Phys. Rev.,' vol. 36, p. 531 (1930).

Finally we wish to express our indebtedness to Professor J. W. Cook, who has supplied us with the materials investigated, with the details of methods of preparation, and throughout has taken the greatest interest in the work. We should like to record our thanks to the International Cancer Research Foundation which has generously financed the research.

SUMMARY

The ultra-violet absorption spectra of ethyl alcohol solutions of a number of aromatic hydrocarbons, some of which are of biological importance, have been investigated. The compounds examined include benzene, naphthalene, anthracene, 1 : 2-benzanthracene, 1 : 2 : 5 : 6-dibenzanthracene, phenanthrene, and their derivatives, as well as chrysene, picene, pyrene and its derivatives.

Details are given of the experimental methods and some discussion of possible errors. The spectra of chemically related compounds are compared and general resemblances between them are pointed out.

In general no adequate theoretical discussion of the absorption spectra of these compounds is possible, but the effects of the addition of various groups have been noticed. Preliminary references are made to changes observed in the absorption spectra on cooling the solutions to approximately $-183°$ C.

Some theoretical considerations of the subject are mentioned.

The Effect of Imprisonment of Resonance Radiation in the Decomposition of Ammonia and of Deutero-Ammonia

By H. W. MELVILLE (Senior Student, Exhibition of 1851), Laboratory of Colloid Science, Cambridge

(*Communicated by Eric K. Rideal, F.R.S.—Received May* 17, 1935)

INTRODUCTION

In three recent papers,* the kinetics of the mercury photo-sensitized decomposition of ammonia and of deutero-ammonia have been compared in order (*a*) to obtain the velocity coefficients for the processes occurring in these reactions, (*b*) to establish correlations between these coefficients and the measured quenching radii and zero-point energy content of the molecules participating in the collisions. Calculations made upon simple premises yielded inconsistencies in the numerical values of the coefficients, thus throwing doubt upon the fundamental assumptions in the kinetic treatment of the problem. Evans and Taylor,† in a later paper, have sought to surmount these difficulties by postulating the formation of mercury-ammonia complexes which have finite and different lifetimes for the two ammonias and which are subject to various deactivating collisions in the gas phase. There are, however, a number of factors to be taken into account, not hitherto considered, before recourse need be made to assumptions of this nature. It is the purpose of this paper to attempt to estimate quantitatively the magnitude of the corrections to be applied, to see then what discrepancies still exist and what probable explanations need be put forward to effect the necessary modifications to the scheme of reactions.

We must distinguish between two separate phenomena occurring in these reactions. In the first place the ammonia molecule is dissociated by a collision with a mercury atom of high energy content and therefore, in order to calculate the velocity of this process, the stationary concentration of the mercury atoms as a function of ammonia and hydrogen pressure and other possible variables must be known. Secondly, the fragments of the dissociated molecule do not necessarily yield molecular hydrogen and nitrogen, for it is known‡ in the direct photo-chemical decomposition the quantum yield is less than unity and therefore reverse actions leading to the regeneration of ammonia must come into play. In comparing the behaviour of isotopic molecules, each of these factors may be different and hence they must be separately investigated. In the following pages the phenomena having to do with mercury atom collisions are described first and then the question of secondary reactions is considered in the second part of the paper.

---

* Jungers and Taylor, 'J. Chem. Phys.,' vol. 2, p. 373 (1934); vol. 2, p. 452 (1934).

† Evans and Taylor, *ibid.*, vol. 2, p. 732 (1934).

‡ Melville, 'Proc. Roy. Soc.,' A, vol. 138, p. 378 (1932).

The two main problems in the kinetic treatment is to account for the variation of reaction rate with ammonia pressure, and the effect of hydrogen, whose function is twofold, (*a*) quenching of excited mercury atoms and (*b*) inhibition of the decomposition by the recombination process $NH_2 + H = NH_3$. Since accurate values of the quenching radii of all the molecules participating in these reactions have been determined,† it is practicable to calculate exactly the kinetic behaviour of the decomposition based on the assumption (*a*) that when a $^3P_1$ Hg atom is deactivated the ammonia molecule is dissociated ($k_1\,[Hg^1]\,[NH_3]$), and (*b*) that H_2 merely quenches without playing any further part in the kinetics ($k_2\,[Hg^1]\,[H_2]$). The rate of equation is

$$1/R = 1/R_\infty \left(1 + \frac{k_3 + k_2\,[H_2]}{k_1\,[NH_3]}\right), \tag{1}$$

where R_∞ is the rate at $[NH_3] = \infty$, pressure broadening being absent and k_3^{-1} the lifetime of the Hg atom. Taking $1/k_3 = 1\cdot07 \times 10^{-7}$ sec* and employing Evans's quenching data† k_1/k_3 is $0\cdot202$ mm^{-1} for NH_3 and $0\cdot073$ for ND_3. For H_2 and D_2 k_2/k_3 is $1\cdot268$. For NH_3 the data of Mitchell and Dickinson‡ yield $k_1/k_3 = 0\cdot20$ mm^{-1}, in good agreement with the calculated value; Jungers and Taylor's results for ND_3 give $0\cdot045$ for k_1/k_3, which is rather smaller than that calculated. On these assumptions the rate of decomposition of the ammonias at high pressures should be identical, for every mercury atom is deactivated by ammonia, and therefore the rate is simply equal to the rate of absorption of light at 2537 A. But this is not so, the ND_3 decomposition being some 10 times faster. Another discrepancy comes to light when the observed and calculated figures are compared for the effect of hydrogen.

In Table I the pressure of hydrogen required to reduce the rate to half its value is computed from the equation

$$k_2\,[H_2] = k_1\,[NH_3] + k_3. \tag{2}$$

---

* Garrett, ' Phys. Rev.,' vol. 40, p. 779 (1932).

† Evans, ' J. Chem. Phys.,' vol. 2, p. 445 (1934).

‡ ' J. Amer. Chem. Soc.,' vol. 49, p. 1478 (1927).

The reactions are much more sensitive to the addition of hydrogen than is to be expected, the calculated values being about 20–40 times greater than the observed. Moreover, this discrepancy cannot be due to inhibition of the decomposition by atomic hydrogen, for if the data on this effect be examined* it will be seen that a much larger proportion of hydrogen must be present to exert a marked effect, *e.g.*, to cut down the rate by one-half it is necessary to have about 0·1 mm H_2 present with 0·3 mm NH_3.

There is, however, another effect which profoundly affects the numerical values of the velocity coefficients as determined from equations (1) and (2). It makes the discrepancies rather larger than those pointed out by Jungers and Taylor and destroys the good agreement between theory and experiment for NH_3. The implicit assumption made in the treatment of problems of this nature is that k_3 or the reciprocal of the mean life of the $^3P_1$ Hg atom is constant. At high mercury-vapour pressures, *e.g.*, 0·001 mm corresponding to a temperature of 15° C, at which experiments are usually made, there is appreciable imprisonment of resonance radiation in tubes of a few centimetres in diameter. For example, if the mean life (τ) be determined by one of the photo-electric methods now employed to measure it very accurately, it is found that τ may increase from $1 \cdot 07 \times 10^{-7}$ sec at a Hg pressure of 10^{-5} mm to as much as 10 times this value—10^{-6} sec—at 0·001 mm.† In performing experiments on the effect of pressure on the rate of decomposition, the value of k_3^{-1} may attain either of the limits. At high pressures where the fraction of Hg atoms deactivated is large $\tau = 1 \cdot 07 \times 10^{-7}$ sec, but at low pressures τ will be 10^{-6} sec in the tubes normally employed. The absolute magnitudes of k_1 are therefore subject to a large uncertainty, although the relative values may not be greatly in error. The treatment of the effect of hydrogen

TABLE I—EFFECT OF HYDROGEN AND DEUTERIUM

$[ND_3]$	$[D_2]_{\frac{1}{2}}$ (obs)	$[D_2]_{\frac{1}{2}}$ (calc)	Calc/obs
40·4	0·060	2·27	37·9
21·3	0·028	1·20	42·9
11·0	0·014	0·62	44·4
4·9	0·011	0·27	24·5
$[NH_3]$	$[H_2]_{\frac{1}{2}}$ (obs)	$[H_2]_{\frac{1}{2}}$ (calc)	Calc/obs
5·06	0·140	1·81	12·9
3·22	0·077	1·44	18·7
1·88	0·055	1·17	21·2
0·96	0·026	0·98	37·8

---

* Melville, 'Trans. Faraday Soc.,' vol. 28, p. 885 (1932).

† Mitchell and Zemansky, "Resonance Radiation," p. 229 (Cambridge).

is likewise subject to just as great an uncertainty, for k_3 comes into equation (2). Without knowing the precise geometry of the apparatus an exact correction is rather difficult to assess, but it is possible on the following lines to apply this correction to the ammonia results.

The variation of the effective mean life of the Hg atom* may be represented to a first approximation by the formula

$$1/\tau - 1/\tau_\infty = 1/\tau_0 \, e^{-kp},$$

where p is the mercury-vapour pressure and k is a constant, or

$$1/\tau = 1/10^{-6} + \frac{1}{1 \cdot 07 \times 10^{-7}} e^{-kp},$$

where τ_∞ is the maximum value of τ attained under any given experimental conditions, and for the present calculations will be taken as 10^{-6} sec. τ_0 is the value of τ at low Hg pressures. The effective mean life would, however, vary in the same way, at high and constant Hg pressure, if foreign gas be added which quenches the resonance radiation, for then a smaller fraction of the atoms radiate and therefore a smaller number of quanta are reabsorbed. Hence

$$1/k_3 = \tau = 1 \cdot 07 \times 10^{-7} + (10 - 1 \cdot 07 \times 10^{-7}) \, e^{-k[\mathrm{NH_3}]}.$$

At low ammonia pressures the value of the second term in the above equation becomes $(10 - 1 \cdot 07) \times 10^{-7}$ and so $\tau = 10^{-6}$ sec, whilst at high pressures the second term may be neglected and τ then amounts to 10^{-6} sec.

It is now necessary to evaluate k. k is connected with the pressure of ammonia $[\mathrm{NH_3}]_{\frac{1}{2}}$, which removes half the atoms excited, by the relationship $\ln 2 = k\,[\mathrm{NH_3}]$ and by $k_1\,[\mathrm{NH_3}]_{\frac{1}{2}} = \tau^{-1}$, where k_1 is the velocity coefficient for deactivation. Consequently

$$k = \ln 2/[\mathrm{NH_3}]_{\frac{1}{2}} = k_1 \tau_0 \ln 2,$$

and hence

$$\tau = 1 \cdot 07 \times 10^{-7} + 9 \cdot 0 \times 10^{-7} \exp\left(- k_1 \tau_0 \ln 2\,[\mathrm{NH_3}]\right).$$

If H_2 be present in addition, the exponent of the exponential term is

$$\tau_0 \ln 2\,(k_1\,[\mathrm{NH_3}] + k_2\,[\mathrm{H_2}]).$$

Therefore, substituting this expression in equation (1),

$$\frac{1}{\mathrm{R}} = \frac{1}{\mathrm{R}_\infty}\left[1 + \frac{k_2[\mathrm{H_2}] + \{1 \cdot 07 \times 10^{-7} + 9 \cdot 0 \times 10^{-7} \exp(-\tau_0 \ln 2\{k_1[\mathrm{NH_3}] + k_2[\mathrm{H_2}]\})\}^{-1}}{k_1\,[\mathrm{NH_3}]}\right],$$

---

* Mitchell and Zemansky, "Resonance Radiation," p. 229 (Cambridge).

and

$$k_2\,[\mathrm{H_2}]_t = k_1\,[\mathrm{NH_3}]$$
$$+\{1\cdot07\times10^{-7}+9\cdot0\times10^{-7}\exp\{-\tau_0\ln 2\,(k_1\,[\mathrm{NH_3}]+k_2\,[\mathrm{H_2}])\}]^{-1}.$$

Expressing times in seconds and pressures in millimetres, $k_1\,(NH_3)$ is $1\cdot88\times10^6$ mm^{-1} sec^{-1}, and $k_1\,(ND_3)$, $0\cdot68\times10^6$ mm^{-1} sec^{-1}; k_2 (H_2 or D_2) is $11\cdot7\times10^6$ mm^{-1} sec^{-1}. Using the above formula, the variation of rate with ammonia and duetero-ammonia pressure has been calculated and compared with Taylor and Jungers's data for ND_3 and Mitchell and Dickinson's for NH_3.

In fig. 1, four curves have been drawn for NH_3 and ND_3. Curve B, fig. 1*b*, is constructed from experiments carried out with high mercury-vapour pressures. C is that calculated from the measured value of the quenching radius and for low Hg pressures, if the mechanism is a simple dissociation $Hg\,(^3P_1) + ND_3 \rightarrow Hg\,(^1S_0) + ND_2 + D$, the agreement not being very good. When correction is made for imprisonment of resonance radiation (curve D) the agreement is very much worse. Even when the experimental conditions are such as to avoid reabsorption (curve A, *see* below) the deviation from curve C is very large. An exactly similar set of curves is obtained with ammonia, fig. 1*a*, except that there is fortuitous agreement between B and C.

The imprisonment of resonance radiation will be less pronounced in narrow tubes and therefore it might be anticipated that the value of $k_1\tau$ would depend on the size of tube. This has already been found for phosphine* (*see* Table II), although at the time the results were obtained it was not realized, but later it was tentatively suggested,† that the inconstancy of the product $k_1\tau$ was due to this cause.

Experimentally the magnitude of $k_1\tau$ is obtained by plotting the reciprocal of the rate of decomposition against the reciprocal of the ammonia pressure. The ratio of the intercept on the 1/R axis to the slope of the line is equal to $k_1\tau$, which is most conveniently expressed in reciprocal millimetres.

It is now necessary to enquire into the nature of the large discrepancy after having corrected for reabsorption.

EXPERIMENTAL

To confirm these calculations, a further examination of the mercury photo-sensitized decompositions of NH_3, ND_3, PH_3, and PD_3 has been undertaken. To avoid reabsorption small mercury-vapour pressures—3×10^{-4} mm and narrow reaction vessels—3 mm—have been employed. In view of the powerful inhibiting effect of hydrogen, a sensitive Pirani

---

* Melville, 'Proc. Roy. Soc.,' A, vol. 138, p. 385 (1932).

† Melville, *ibid.*, vol. 139, p. 547 (1933).

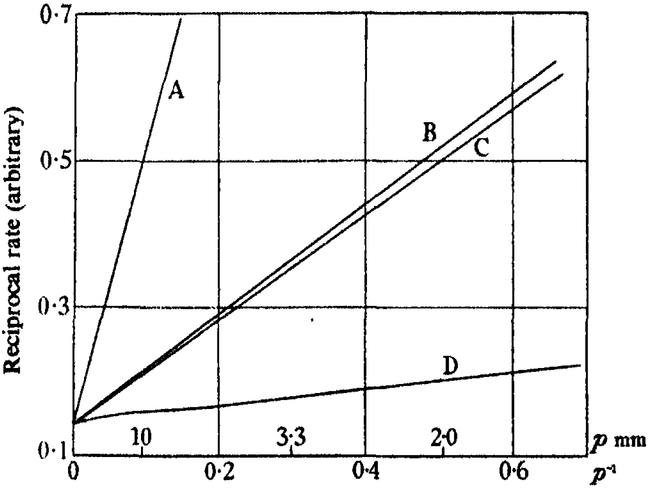


FIG. 1a—NH_3.

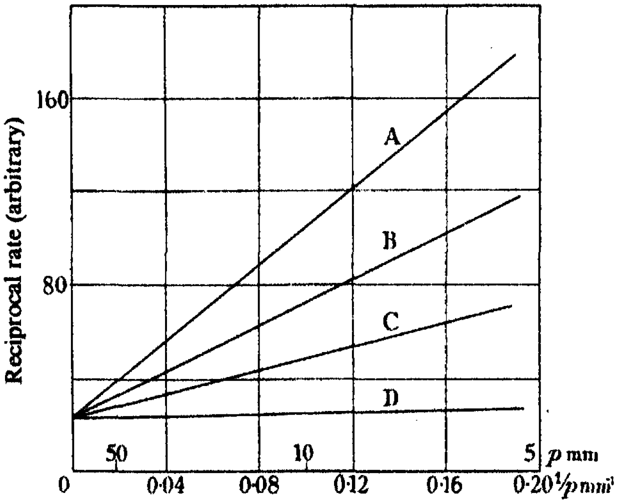


FIG. 1b—ND_3.

gauge of small volume was used to measure the pressure of the hydrogen and nitrogen after the ammonia had been condensed out with liquid air. The initial rates of decomposition could thus be measured fairly accurately. The mercury lamp was of the water-cooled type and a 3-cm layer of 50% acetic acid served to eliminate the direct dissociation of these molecules.

Ammonia was prepared from soda lime and ammonium chloride, tri-deutero-ammonia from magnesium nitride (by passing nitrogen over magnesium) and 97% heavy water and phosphine from PH_4I and KOH. Tri-deutero-phosphine was made by the interaction of 99% heavy water and calcium phosphide. After preparation its deuterium content was found to be 98%. All these gases were subjected to fractionation *in vacuo* until absolutely free from gases non-condensable in liquid air.

TABLE II

Gas	$k_1\tau$ (mm^{-1})	Conditions	Pressure range (mm)
NH_3	0·20	Mitchell and Dickinson, high Hg pressure, wide tube	1–10
	0·10	Jungers and Taylor, high Hg pressure, wide tube	4–50
	0·029	Melville, present experiments, low Hg pressure, narrow tube	5–500
	0·20	Theory if mechanism is $Hg' + NH_3 \rightarrow NH_2 + H + Hg$	—
ND_3	0·045	Jungers and Taylor, high Hg pressure, wide tube.	5–40
	0·025	Melville, present experiments, low Hg pressure, narrow tube	5–500
	0·073	Theory if mechanism is $Hg' + ND_3 \rightarrow ND_2 + H + Hg$	—
PH_3	2·7	Melville, high Hg pressure, 2 cm tube............	—
	1·6	Melville, high Hg pressure, 1 cm tube............	—
	1·15	Melville, high Hg pressure, 0·5 cm tube..........	—
	0·256*	Low Hg pressure, narrow tube	—
PD_3	0·318*	Low Hg pressure, narrow tube.................	—

* Unpublished observations.

All the available data, including those obtained during the present experiments, on the value of $k_1\tau$ have been collected in Table II, which also indicates the conditions under which this magnitude was obtained. In every example it will be observed that the value of $k_1\tau$ is diminished when the reabsorption is practically completely eliminated and the effective mean life of the atom is 10^{-7} sec. With PH_3 the effect of tube diameter must therefore be wholly ascribed to this effect and not to any other processes connected with secondary reactions on the walls of the vessel as was first suggested. As has been mentioned, this correction to the NH_3 and ND_3 results renders the agreement between theoretical and experimental values much worse than that previously calculated, which suggests that the reaction mechanism is more complicated than that hitherto supposed.

In performing experiments of this nature, the pressure range has always been restricted in order to avoid pressure broadening of the absorption line at 2537 A. To find how far this factor really does affect the velocity, pressures up to atmospheric were employed in determining the value of $k_1\tau$ for NH_3 and for ND_3. No experimentally measurable deviation from the simple formula $1/R - 1/[NH_3]$ could be detected, so that pressure broadening may be neglected for the present purpose.

The interpretation of the relative and absolute values of the quenching radii of NH_3 and of ND_3 yields strong evidence in favour of the argument that these molecules deactivate the excited atoms to the meta-stable $^3P_0$ level only. The energy absorbed by the ammonia molecule in this collision is so small that it cannot lead to dissociation. It was pointed out on p. 327 that the inhibitory effect of hydrogen and deuterium was many times greater than that calculated from the known quenching radii of the molecules. The simplest way to explain these apparently anomalous results is to postulate that the reaction responsible for the decomposition of the ammonia is

$$\mathrm{Hg}\,(^3P_0) + NH_3 \rightarrow \mathrm{Hg}\,(^1S_0) + NH_2 + H.$$

By postponing the dissociation to this stage, the hydrogen molecules, which are very efficient in deactivating metastable atoms,* now have a much more favourable opportunity of exerting a large influence on the rate of dissociation of ammonia. This suggestion also accounts for the difference between the value of k_1 calculated from the quenching data and k_1 calculated from the new data based on the simpler theory.

The modified reaction scheme is depicted below, the previous scheme being shown for comparison

$$\mathrm{Hg}\,(^3P_1)\begin{cases} \xrightarrow{k_3\,\text{(radiate)}} \mathrm{Hg}\,(^1S_0) \\ \xrightarrow{k_1\,(NH_3)} \mathrm{Hg}\,(^1S_0) + NH_2 + H \\ \xrightarrow{k_2\,(H_2)} \mathrm{Hg}\,(^1S_0) + 2H \end{cases}$$

Old Scheme

$$\begin{tikzcd}[column sep=large]
 & \mathrm{Hg}\,(^1S_0) & \\
\mathrm{Hg}\,(^3P_1) \arrow[ur, "k_3\,\text{(radiate)}"] \arrow[r, shift left, "k_1\,(NH_3)"] & \mathrm{Hg}\,(^3P_0) \arrow[l, shift left, "k'_1\,(NH_3)"] \arrow[d, "k_4\,(NH_3)"] \arrow[dd, bend right, "k_5\,(H_2)"'] & \\
 & \mathrm{Hg}\,(^1S_0) + NH_2 + H & \\
 & \mathrm{Hg}\,(^1S_0) + 2H &
\end{tikzcd}$$

New Scheme

---

* *See, e.g.*, Klumb and Pringsheim, 'Z. Physik,' vol. 52, p. 610 (1929).

In the new scheme a simplification has been made by omitting direct deactivation of $^3P_1$ atoms by H_2 since in nearly all the experiments this is negligible compared with the rate of deactivation by ammonia. One additional process has been added, namely, the raising of the $^3P_0$ atoms to the $^3P_1$ level by collisions with ammonia molecules of sufficiently high energy content, a process which has been shown to occur and has been discussed in detail by Gaviola.*

To evaluate the coefficients k_4 and k_5, the normal procedure is adopted. First it will be supposed that hydrogen is absent. Therefore the stationary state equations for excited and metastable atoms are

$$d[^3P_1]/dt = K + k_1'[^3P_0][NH_3] - k_3[^3P_1] - k_1[^3P_1][NH_3] = 0$$
$$d[^3P_0]/dt = k_1[^3P_1][NH_3] - k_1'[^3P_0][NH_3] - k_4[^3P_0][NH_3] = 0,$$

where K is the rate at which $^3P_1$ atoms are produced by photo excitation. $[^3P_1]$ and $[^3P_0]$ refer to the concentration of mercury atoms in these levels. Solving

$$[^3P_1] = \frac{K(1 + k'_1/k_4)}{k'_1k_3/k_4 + k_3 + k_1[NH_3]}$$

$$[^3P_0] = \frac{K}{k_4[NH_3]}\left\{1 - \frac{(1 + k'_1/k_4)k_3}{k'_1k_3/k_4 + k_3 + k_1[NH_3]}\right\},$$

and therefore

$$R = -\frac{d[NH_3]}{dt} = \frac{\alpha K k_1[NH_3]}{k_1[NH_3] + (1 + k'_1/k_4)k_3},$$

or

$$1/R = \frac{1}{\alpha K}\left\{1 + \frac{(1 + k'_1/k_4)k_3}{k_1[NH_3]}\right\},$$

where α is a coefficient expressing how many molecules of NH_3, originally dissociated, ultimately form N_2 and H_2. This is probably less than unity, owing to reverse reactions such as $NH_2 + H$, unconnected with mercury atom processes. The ratio of the intercept to slope of the $1/R - 1/[NH_3]$ plot, which is linear and therefore in accordance with experiment, now becomes

$$\frac{k_1}{k_3(1 + k'_1/k_4)} \quad \text{or} \quad k_1\tau/(1 + k'_1/k_4)$$

compared with $k_1\tau$ on the simpler theory. $k_1\tau$ is known from quenching data and therefore k'_1/k_4 may be evaluated. This ratio is 5·9 for NH_3 and 1·9 for ND_3. k'_1 is proportional to $\sigma^2e^{-E/RT}$ where E is the difference in energy between the $^3P_1$ and $^3P_0$ levels, *i.e.*, 0·218 volts or 5000 cal and

---

* 'Phys. Rev.,' vol. 34, p. 1373 (1929).

σ is the sum of the radii of the mercury atom and the ammonia molecule, and, as pointed out by Gaviola, is equal numerically to that derived from quenching experiments and not to that derived from kinetic theory. The ratio

$$\frac{k_4\,(\mathrm{ND_3})}{k_4\,(\mathrm{NH_3})}$$

is therefore

$$\frac{5 \cdot 9}{1 \cdot 9} \times \frac{0 \cdot 073^*}{0 \cdot 20} = 1 \cdot 13,$$

that is, deutero-ammonia reacts slightly more quickly than ammonia, this behaviour being paralleled by H_2 and D_2, D_2 having the larger quenching radius.

Assuming that the collision radii for NH_3 and ND_3 with $^3P_0$ atoms in the dissociation process (k_4) are identical, then

$$\frac{k'_1}{k_4} = \frac{\sigma_Q{}^2\, e^{-5000/RT}}{\sigma_{K.T.}{}^2\, e^{-x/RT}},$$

where $\sigma_{K.T.}$ is the kinetic theory value of the sum of the radii of the atom and molecule and x is the energy of activation for the reaction whose velocity coefficient is k_4. From this equation

$$x_{NH_3} = 6700 \text{ cal} \qquad \text{and} \qquad x_{ND_3} = 5700.$$

As was mentioned on p. 327, the effect of hydrogen is twofold—deactivation of excited atoms and inhibition by atomic hydrogen. The first effect may now be considered. The only modification of the above equations is the introduction of a term expressing the fact that $^3P_0$ atoms are now removed by H_2 molecules at a rate $k_5\,[^3P_0]\,[H_2]$. Hence

$$-\frac{d\,[\mathrm{NH_3}]}{dt} = \alpha k_4\,[^3P_0]\,[\mathrm{NH_3}]$$

$$= \frac{\alpha K k_1 k_4\,[\mathrm{NH_3}]^2/k_3}{k_1 k_4\,[\mathrm{NH_3}]^2/k_3 + k_1 k_5\,[\mathrm{H_2}]\,[\mathrm{NH_3}]/k_3 + k'_1\,[\mathrm{NH_3}] + k_4\,[\mathrm{NH_3}] + k_5\,[\mathrm{H_2}]}.$$

In absence of H_2

$$-\frac{d\,[\mathrm{NH_3}]}{dt} = \frac{\alpha K k_1 k_4\,[\mathrm{NH_3}]^2/k_3}{k_1 k_4\,[\mathrm{NH_3}]^2/k_3 + k'_1\,[\mathrm{NH_3}] + k_4\,[\mathrm{NH_3}]}.$$

If $[H_2]_{\frac{1}{2}}$ is the hydrogen concentration required to reduce the rate of half its value in absence of hydrogen then

$$k_1 k_4\,[\mathrm{NH_3}]^2/k_3 + (k'_1 + k_4)\,[\mathrm{NH_3}] = \{k_1/k_3 \,.\, [\mathrm{NH_3}] + 1\}\, k_5\,[\mathrm{H_2}]_{\frac{1}{2}}.$$

---

* See p. 326.

By inserting the appropriate values of k_1/k_3 for NH_3 and ND_3

$$k_5/k_4 \,.\, [H_2]_t = [NH_3]\left\{\frac{0{\cdot}20\,[NH_3] + 6{\cdot}9}{0{\cdot}20\,[NH_3] + 1{\cdot}0}\right\}, \qquad (A)$$

and

$$k_5\,/k_4 \,.\, [D_2]_t = [ND_3]\left\{\frac{0{\cdot}073\,[ND_3] + 2{\cdot}9}{0{\cdot}073\,[ND_3] + 1{\cdot}0}\right\}, \qquad (B)$$

the following results were obtained (Table III). First it will be seen that D_2 is a more efficient inhibitor for the ND_3 decomposition than H_2 is for the NH_3 reaction. From the data in Table III

$$k_4/k_5 \doteqdot 5 \times 10^{-4}\,(NH_3), \doteqdot 5 \times 10^{-4}\,(ND_3).$$

If the collisions between H_2 and $^3P_0$ atoms are of the same order of efficiency as those between H_2 and $^3P_1$ atoms, 5×10^{-4} is the probability of the reaction of Hg ($^3P_0$) with NH_3; this corresponds to an energy of activation of 4200 cal in reasonable agreement with the value calculated by the more accurate method described on p. 334.

TABLE III

$[NH_3]$	$[NH_3]\left\{\frac{0{\cdot}20[NH_3]+6{\cdot}9}{0{\cdot}20[NH_3]+1{\cdot}0}\right\}$	$[H_2]_t$	$[ND_3]$	$[ND_3]\left\{\frac{0{\cdot}073[ND_3]+2{\cdot}9}{0{\cdot}073[ND_3]+1{\cdot}0}\right\}$	$[D_2]_t$
26	51	0·030	27	44	0·011
17	40	0·024	—	—	—
10	30	0·014	10·6	22	0·0075
5·5	21	0·008	5·4	13	0·0055
1·5	8·3	0·003	—	—	—

It has been assumed in the calculations in the preceding paragraph that k_5 for H_2 and D_2 are identical. This was verified by determining the influence of D_2 on the NH_3 reaction and of H_2 on the ND_3 reaction. First a run was made with NH_3, curve 1, fig. 2*a*, second a similar run with D_2 present at the start, curve 2. Curve 2 has been displaced to the right until its origin was on curve 1, thus giving curve 3. The fact that curves 2 and 3 lie so close together shows that D_2 exerts the same influence as H_2 in inhibiting the decomposition of NH_3 in this pressure range. The corresponding experiments with ND_3 and H_2 yielded similar results, fig. 2*b*.

THE ABSOLUTE RATES OF THE REACTIONS

At sufficiently high pressures every mercury atom excited dissociates an ammonia molecule, and therefore if α is the same for NH_3 as it is for ND_3 the relative rates of decomposition should also be identical. Deutero-ammonia decomposes under these conditions about 10 times more slowly than ammonia, Taylor and Jungers's data yield 1 : 10, while the present experiments yield ratios which varied a little, one series giving 1 : 9·7 and another 1 : 12·4. If the foregoing theory is correct, this difference would be attributed to the fact that the secondary reactions after the primary dissociation of the NH_3 molecule are 10 times more

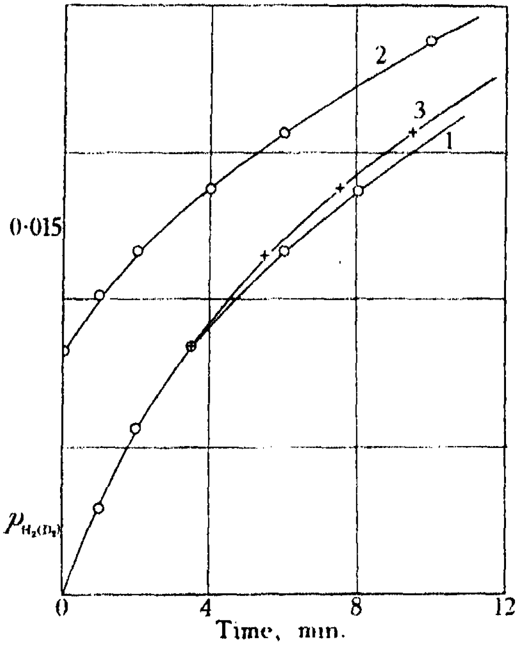


FIG. 2a—D_2 on NH_3.

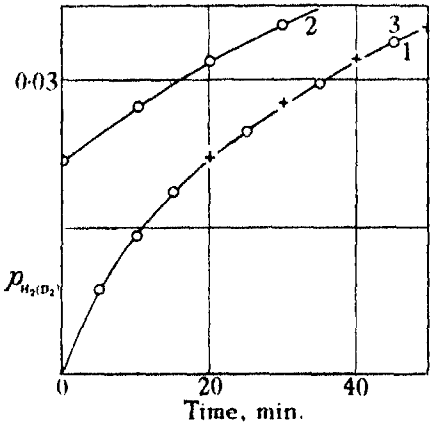


FIG. 2b—H_2 on ND_3.

effective than those in the dissociation of ND_3. Two methods exist for confirming this supposition: (*a*) comparison of the direct decomposition rates, and (*b*) measurement of the extent of inhibition of the direct photodissociation of ND_3 by atomic deuterium.

The reaction vessel used in the photo-sensitized experiments was replaced by a cylindrical vessel 2 cm in diameter. The acetic acid filter was replaced by one of distilled water and mercury vapour kept out of the tube by a trap cooled to −80°. The effectiveness of the trap was tested by interposing the acetic acid filter when the rate of decomposition was reduced to less than 1% of its previous value (*cf.* Table IV). At 100 mm pressure of NH_3, 96% of the light at 2100 A is absorbed in a layer 2 cm thick, and therefore if ND_3 has a slightly different absorption coefficient the results, on this account alone, will not differ by more than a few units per cent. The temperature of the mercury trap was allowed to rise a little in order to keep the ammonia pressure at 100 mm, since the vapour pressure of NH_3 and of ND_3 at −80° is less than 100 mm. Three series of experiments yielded relative rates $NH_3 : ND_3$ 2·6, 2·3, and 2·6 : 1, while the relative rates of the photo-sensitized reactions at 100 mm was 7 : 1, which is somewhat less than in the narrow tubes employed for the majority of the above experiments. Although these ratios are not identical, yet they demonstrate that the large ratio of rates in the mercury-sensitized reactions is due partly, if not wholly, to the intervention of secondary reactions having nothing to do with collisions with excited mercury atoms.

INHIBITION BY HYDROGEN AND BY DEUTERIUM ATOMS

Since the secondary reactions in the decomposition of deutero-ammonia are less effective than those of ammonia, it would be expected that the dissociation of this molecule would be more easily inhibited by atomic deuterium than that of ammonia by atomic hydrogen. Here again further information may be gleaned by studying the inhibition of the decomposition of NH_3 by atomic D and of NH_3 by atomic H.

Instead of employing a zinc spark for the direct reaction, as was done originally, the mercury lamp was so arranged to provide the light required for both the direct and sensitized dissociation. The procedure was as follows. The mixture of ammonia and hydrogen was introduced into the reaction vessel and the rate of the direct decomposition measured, the mercury-vapour trap being at −80°. Distilled water filled the cell normally used for the acetic acid filter. The mercury trap was then heated to 0° C and the rate again measured; this gives the velocity of

TABLE IV—INHIBITION BY ATOMIC HYDROGEN AND DEUTERIUM

Expt. No.	Time of exposure	Total pressure	Increase of pressure	Conditions
NH$_3$ 3·35 mm, H$_2$ added				
1	0	0·0635	—	—
	5	0·0782, 0·0805*	0·0147	Direct (D)
	10	0·0897	0·0092	Direct + Sensitized (S)
	15	0·0890	0·0007	Sensitized
NH$_3$ 3·09 mm, 70% D added				
2	0	0·0638	—	—
	5	0·0933, 0·0937*	0·0295	D
	10	0·1118	0·0161	D + S
	15	0·1108	0·0010	S
ND$_3$ 3·18 mm, H$_2$ added				
3	0	0·0677	—	—
	20	0·0908, 0·0913*	0·0231	D
	40	0·0889	0·0024	D + S
	60	0·0849	0·0040	S
ND$_3$ 3·18 mm, D$_2$ added				
4	0	0·0625	—	—
	10	0·0793, 0·0823*	0·0168	D
	20	0·0803	0·0020	D + S
	30	0·0753	0·0053	S
ND$_3$ 10·3 mm, H$_2$ added				
5	0	0·1054	—	—
	5	0·1206, 0·1240*	0·0152	D
	10	0·1322	0·0082	D + S
	15	0·1310	0·0012	S
ND$_3$ 10·3 mm, 70% D added				
6	0	0·0923	—	—
	5	0·1132, 0·1166*	0·0209	D
	10	0·1276	0·0110	D + S
	15	0·1260	0·0016	S

* This pressure increase is due to warming the mercury vapour trap from −80° to 0°.

the direct and sensitized reactions together in presence of atomic hydrogen. Finally the acetic acid filter was interposed and the rate of the sensitized reaction alone determined. In the following experiments the pressure of hydrogen or of deuterium was made sufficiently high to inhibit the sensitized reaction almost completely, and therefore the difference in rate between the first two experiments shows directly the extent of inhibition of the direct reaction by atomic hydrogen.

The negative pressure increase is due to the clean-up of a small part of the atomic hydrogen or deuterium. Experiments 1 and 2 demonstrate that deuterium exerts the same inhibitory influence as hydrogen. Comparing experiments 1 and 2 with 3 and 4 shows that the decomposition of deutero-ammonia is much more susceptible to inhibition than ammonia, the direct reaction being almost completely suppressed. This makes it difficult to determine whether H_2 and D_2 inhibit equally well, and therefore in experiments 5 and 6 the ND_3 pressure has been increased so that the extent of the inhibition may be measured. As with NH_3, hydrogen and deuterium atoms are equally efficient inhibitors.

These experiments now enable a more detailed analysis to be made of the decomposition of NH_3 and of ND_3, whereby the secondary reactions responsible for the greater reactivity of NH_3 may be more precisely specified.

In suggesting a simple tentative mechanism for the dissociation of these molecules, one important fact must be borne in mind, namely, that the quantum yield for the dissociation of ammonia is practically independent of pressure between 1 and 10^4 mm.* The simplest way to interpret this fact in evolving a scheme of reactions is to suppose that all secondary reactions are either unaffected or affected in the same way by pressure. They may occur in the gas or on the walls. The equations for the stationary concentrations of H and of NH_2 therefore simplify to

$$\frac{d\,[\mathrm{H}]}{dt} = \mathrm{I}\,[\mathrm{NH_3}] - k_1\,[\mathrm{H}]^2 - k_2\,[\mathrm{H}]\,[\mathrm{NH_2}] = 0$$

$$\frac{d\,[\mathrm{NH_2}]}{dt} = \mathrm{I}\,[\mathrm{NH_3}] - k_2\,[\mathrm{H}]\,[\mathrm{NH_2}] - k_3\,[\mathrm{NH_2}]^2 = 0,$$

presuming the reactions to be $H + H \rightarrow H_2\,(k_1)$, $H + NH_2 \rightarrow NH_3\,(k_2)$, and $NH_2 + NH_2 \rightarrow N_2 + 2H_2\,(k_3)$. Therefore

$$-\frac{d\,[\mathrm{NH_3}]}{dt} = \mathrm{I}\,([\mathrm{NH_3}] - k_2\,[\mathrm{H}]\,[\mathrm{NH_2}]$$

$$= \mathrm{I}\,[\mathrm{NH_3}]\left(1 - \frac{k_2}{\sqrt{k_1 k_3} + k_2}\right).$$

---

* Wiig and Kistiakowski, 'J. Amer. Chem. Soc.,' vol. 54, p. 1806 (1932); Ogg, Leighton, and Bergstrom, *ibid.*, vol. 56, p. 318 (1934).

The quantum yield is given by

$$\gamma = \frac{\sqrt{k_1k_3}}{\sqrt{k_1k_3} + k_2}$$

and is independent of I and of [NH_3], which is in agreement with experiment. But

$$\gamma_{NH_3} = 0\cdot25$$

and hence

$$k_2/(k_1k_3)^{\frac{1}{2}} = 3$$

and

$$\gamma_{ND_3} = 0\cdot10$$

and hence

$$k_2/(k_1k_3)^{\frac{1}{2}} = 9.$$

Since the efficiency of H and of D atoms in inhibiting the decomposition of NH_3 and of ND_3 is equal, it follows that the ratio k_2/k_1 is not altered by substituting D for H. The greater quantum yield in the NH_3 reaction is therefore mainly due to $k_3\,(NH_2)$ being larger than $k_3\,(ND_2)$. This also explains why the decomposition of ND_3 is more strongly inhibited by atomic hydrogen or deuterium. For a given intensity of illumination, deutero-ammonia produces a higher stationary concentration of ND_2 radicals than ammonia does of NH_2 radicals, and therefore D or H atoms obtain a better chance to participate in the reverse reaction

$$D\,(H) + ND_2 \rightarrow ND_3\,(ND_2H).$$

The greater reactivity of the NH_2 radical is in accordance with previous experience in kinetics, namely, that hydrides react more rapidly than deuterides. The ratio $k_3\,(NH_2)/k_3\,(ND)_2$ is 9 and, if the collision radii are the same, this would correspond to a difference in energy of activation of 1280 cals.

The author is greatly indebted to Professor E. K. Rideal, F.R.S., for much advice and encouragement throughout the progress of this work. He also desires to thank the Royal Commissioners for the Exhibition of 1851 for a senior Studentship.

SUMMARY

Velocity coefficients in mercury photo-sensitized reaction are usually determined by comparison with the mean life of the excited $^3P_1$ atom which is presumed to be constant. At high mercury-vapour pressures—0·001 mm—at which the majority of reactions have been studied, there

is appreciable reabsorption of resonance radiation with the result that the effective mean life is much larger than its actual value of 10^{-7} sec, and may vary, to a large extent depending on the experimental conditions. Absolute and even relative values of velocity coefficients are therefore subject to much uncertainty.

A re-examination of the decomposition of NH_3, ND_3, PH_3, and PD_3 has been undertaken and it is shown, by operating under conditions where reabsorption is negligible, that smaller velocity coefficients are obtained compared with those cited in previous publications, in agreement with expectation.

This leads to a new interpretation of the behaviour of NH_3 and ND_3 molecules. It is suggested that the dissociation is brought about in collisions between NH_3 (ND_3) and metastable atoms, derived in turn from quenching collisions between $^3P_1$ atoms and NH_3 (ND_3). It is shown that the former collisions are almost equally efficient for NH_3 and for ND_3. This mechanism also explains the abnormally high inhibitory effect of hydrogen which is 20–40 times greater than that computed from the relative quenching radii of NH_3 and H_2 and ND_3 and D_2.

To explain why NH_3 decomposes more quickly than ND_3 it is necessary to assume that secondary reactions, unconnected with mercury atom processes, are responsible. Confirmation of this view is obtained from the observation that in the direct photo-decomposition NH_3 also reacts more quickly. Further evidence supporting this contention comes from the fact that the decomposition of ND_3 is more susceptible to inhibition by atomic deuterium than NH_3 by atomic hydrogen.

The Diamagnetic Susceptibilities of Salts Forming Ions with Inert Gas Configurations

II—The Alkaline Halides

By G. W. Brindley, Ph.D., Assistant Lecturer in the University of Leeds, and F. E. Hoare, Ph.D., Lecturer in Physics in the University College, Exeter

(*Communicated by R. Whiddington, F.R.S.—Received May* 24, 1935)

1—Introduction

The data so far published on the diamagnetic susceptibilities of the alkaline halides, measured for the salts in the crystalline state, are very discordant and incomplete, as reference to Table I will show. The aim in carrying out these experiments has been twofold: firstly, to obtain a complete set of values for these salts and secondly to examine more closely than has hitherto been possible how rigorously the susceptibilities of simple crystalline salts are additive. It has already been established that the susceptibilities are approximately additive, but it has not been possible to test this with exactitude because of (*a*) the large discrepancies between the results obtained by previous observers, and (*b*) the lack of data for many crystals. The discrepancies may have arisen to some extent from the different experimental methods, some of which are more accurate than others and some of which may introduce errors peculiar to themselves. We have therefore made a complete re-determination of the susceptibilities of all the alkaline halides, using the same apparatus and method under the same conditions. Since any systematic experimental errors will affect all our results to approximately the same extent, we shall be in a stronger position for testing the additivity of the susceptibilities than if we rely partly on our own and partly on other observers' results. Previous investigators have measured the susceptibilities of some compounds in the crystalline state and others in solution; the latter are of no help in connexion with our problem, for an examination of the available data suggests that solutions have susceptibilities higher by several per cent. than the corresponding crystals. We cannot, therefore, arrive at any certain conclusion by using results obtained partly for crystals and partly for solutions.

TABLE I—MOLECULAR DIAMAGNETIC SUSCEPTIBILITIES × 10^6 OF THE ALKALINES HALIDES IN THE CRYSTALLINE STATE

Salt	(i) Present results (ii) Number of samples used		Other observers		
	(i)	(ii)			
LiF	$10{\cdot}1_4 \pm 0{\cdot}04$	4			
LiCl	23·3 ± 0·11	4			
LiBr	34·0 ± 0·16	4			
LiI	50 (?)				
NaF	15·6 ± 0·19	10		19·6†	
NaCl	30·2 ± 0·14	4	30·1*	29·6†	29·2¶
NaBr	41·1 ± 0·23	4			
NaI	57·0 ± 0·23	4			
KF	23·6 ± 0·06	4		25·7†	
KCl	38·8 ± 0·09	4	39·1*	35·8†	36·15§
KBr	49·2 ± 0·17	4	49·15*	48·2†	48·1§
KI	65·7 ± 0·22	4	63·8*	62·7†	63·9§
RbF	31·9 ± 0·16	2			
RbCl	46·4 ± 0·21	3			45·0‖
RbBr	56·4 ± 0·21	3		62·9‡	56·7‖
RbI	72·2 ± 0·62	3			67·0‖
CsF	44·5 ± 0·21	2			
CsCl	56·7 ± 0·19	3			
CsBr	67·2 ± 0·26	3			
CsI	82·6 ± 0·18	3		77·6‡	

* Hocart, ' C.R. Acad. Sci., Paris,' vol. 188, p. 1151 (1929).
† Kido, ' Sci. Rep. Imp. Univ. Tôhoku,' vol. 21, pp. 149, 288 (1932).
‡ Crow, ' Trans. Roy. Soc. Canada,' vol. 19, p. 63 (1925).
§ Gray and Farquharson, ' Phil. Mag.,' vol. 10, p. 191 (1930).
‖ Gray and Dakers, ' Phil. Mag.,' vol. 11, p. 81 (1931).
¶ Ishiwara, ' Sci. Rep. Imp. Univ. Tôhoku,' vol. 3, p. 303 (1914).

2—METHOD

The method previously described* has been used to obtain a complete set of values for the susceptibilities of the alkaline halides. Although slight modifications have been made from time to time, the apparatus has remained, in essentials, the same as when used for the measurement of the susceptibilities of the sodium and potassium halides.

---

* Hoare, ' Proc. Roy. Soc.,' A, vol. 147, p. 88 (1934). A misprint occurs on p. 93 of this paper. The value of K_A should have been given as $+ 29{\cdot}0 \times 10^{-6}$ c.g.s. The correct value was used in the calculations.

All the salts used in this investigation were specially prepared and purified by British Drug Houses, Ltd.; an additional specimen of caesium iodide was obtained from Griffin & Tatlock. Before use, all the salts were carefully dried in a platinum dish. Measurements were made upon at least two entirely distinct samples of each salt, so that it is unlikely that errors have arisen from the salts being imperfectly packed in the specimen tubes.

For the lithium salts, the same specimen tubes were used as in the work previously described, but for the more expensive salts of rubidium and caesium, of which strictly limited quantities were available, tubes of smaller capacity were used. This enabled several packings of each salt to be employed without repeatedly using the same sample, and hence avoided possible contamination in extracting the salt from the tube and repacking it. Measurements for certain salts were made with two tubes, and their satisfactory agreement is sufficient guarantee that no systematic errors were introduced by the use of but a single tube in other cases.

3—DISCUSSION OF RESULTS

The results obtained in the course of the present investigation are given in the second column of Table I; the ± sign indicates the average deviation from the mean. This column also gives the number of samples of each salt used for which entirely independent results were obtained; two measurements were made upon each sample. The results for caesium iodide from the two different sources were in excellent agreement.

The result for lithium iodide calls for special mention. This salt is difficult to prepare in the anhydrous state, and for these measurements only a hydrate was available. Taking the results obtained with this salt, and assuming the susceptibility of the water to be additive, a value of approximately 50×10^{-6} was obtained for the anhydrous salt. This figure has been adopted provisionally, but it is hoped that a specimen of the pure anhydrous salt will be available in the near future, when this result can be revised if necessary.

In Table I, the most recent results of other investigators for the susceptibilities of salts in the crystalline state are also given. The excellent agreement of three of Hocart's results with those of the present investigation can be taken as establishing the diamagnetic susceptibilities of these salts to an accuracy exceeding 1%, since Hocart confined his attention to an exact determination of the susceptibilities of a small number of crystalline salts. The general trend of the results given in the table indicates that Hocart's value for potassium iodide is probably low.

4—THE ADDITIVITY OF DIAMAGNETIC SUSCEPTIBILITIES

The simplest way to test the additive law of susceptibilities is to tabulate the results as in Table II and then take horizontal and vertical differences, Δ; these should be constant in horizontal and vertical lines if the susceptibilities are strictly additive. They are clearly not constant and therefore the susceptibilities are not strictly additive. A little consideration shows, however, that small experimental errors may be considerably overemphasized by this manner of presentation; consider, for example, the Δ values for the four salts NaCl, NaBr, KCl, and KBr, which are given in Table II as 10·9, 10·4, 8·6, and 8·1. If the four values of χ are in error by 0·1, the NaCl and KBr values *low* by this amount and the NaBr and KCl values *high* by the same amount, then the corrected Δ values become 10·7, 10·6. 8·4, and 8·3, *i.e.*, the two pairs of values become much more nearly constant. This example shows that errors as small as 0·1 in χ (which are of the order of ½%), distributed in an unfavourable way, may produce variations in the Δ values of 0·4, *i.e.*, four times as great as the original errors.

TABLE II—ADDITIVITY OF SUSCEPTIBILITIES

	F	Δ	Cl	Δ	Br	Δ	I
Li	10·14	13·2	23·3	10·7	34·0	16·0	50·0
Δ	5·5		6·9		7·1		7·0
Na........	15·6	14·6	30·2	10·9	41·1	15·9	57·0
Δ	8·0		8·6		8·1		8·7
K	23·6	15·2	38·8	10·4	49·2	16·5	65·7
Δ	8·3		7·6		7·2		6·5
Rb........	31·9	14·5	46·4	10·0	56·4	15·8	72·2
Δ	12·6		10·3		10·8		10·4
Cs	44·5	12·2	56·7	10·5	67·2	15·4	82·6

Apart from the smaller variations of Δ, however, there remains a number of larger variations which are unlikely to have arisen from experimental errors; for example, the differences between CsCl and CsF and between LiCl and LiF are considerably smaller than between other chlorides and fluorides. The difference between RbF and CsF is greater than between the other Rb and Cs salts.

An explanation of these variations in the Δ values is to be sought in terms of the effects which the ions have on each other when they are packed together in a crystal lattice. Useful clues may be obtained from a consideration of other physical properties which are approximately additive, *e.g.*, ionic packing sizes, or properties which might be expected to show a steady progression in passing from LiF at one extreme to CsI at the other, *e.g.*, cohesive properties such as the melting-points.

Studies of interatomic distances in crystals by Goldschmidt,* Pauling,† Zachariasen,‡ and others have shown that in many crystals ionic packing sizes depend mainly on the magnitude of the Coulomb forces between adjacent ions and on the co-ordination number (the number of ions of one kind surrounding one of the opposite kind). Zachariasen, following similar methods to Pauling, has tabulated the packing sizes of ions, assuming them to be singly charged and having a co-ordination number 6, and has also tabulated the corrections to be applied for other charges and co-ordination numbers. Table III gives the packing radii of singly charged ions with co-ordination numbers 6 and 8, and compares the calculated and observed interatomic distances in the alkaline halides. The crystal structures of these salts are all of the rock salt type for which the co-ordination number is 6, with the exception of CsCl, CsBr, and CsI, which have the so-called caesium chloride type of structure with a co-ordination number 8. The good agreement between the observed and calculated interatomic distances indicates (*a*) that the packing sizes in most of these crystals are closely additive, and (*b*) that Zachariasen's corrections of the packing sizes from co-ordination number 6 to 8 are fairly accurate.

The only crystals for which there are large discrepancies between the observed and calculated interatomic distances are LiCl, LiBr, LiI, NaBr (?), and NaI. Pauling has suggested the following explanation. According to Born's theory of ionic lattices, the principal forces between the ions are the Coulomb forces of attraction and repulsion, and the intrinsic repulsions between adjacent ions due to the overlapping of their wave functions. Since the latter vary as a high inverse power of the distance separating the ions, they are important only between adjacent ions which, in most of the alkaline halides, are pairs of positive and negative ions. But in a few crystals like LiI, owing to the small size of the positive ions, the negative ions are brought into specially close proximity and in consequence the intrinsic repulsions between pairs of

---

* 'Geochemische Verteilungsgesetze,' Oslo (1927); 'Trans. Faraday Soc.,' vol. 25, p. 253 (1929).

† 'Z. Kristallog.,' vol. 67, p. 377 (1928); 'J. Amer. Chem. Soc.,' vol. 49, p. 763 (1927).

‡ 'Z. Kristallog.,' vol. 80, p. 137 (1931).

TABLE III

(i) Ionic packing radii, in A (After Zachariasen)

	Li^+	Na^+	K^+	Rb^+	Cs^+
Co-ordination number 6 ..	0·68	0·98	1·33	1·48	1·67
„ „ 8 ..	—	—	—	—	1·71
		F^-	Cl^-	Br^-	I^-
Co-ordination number 6 ..		1·33	1·81	1·96	2·19
„ „ 8 ..		1·39	1·87	2·01	2·24

(ii) Observed and calculated interatomic distances, *r*, in A

$\Delta r = r_{\text{Obs.}} - r_{\text{Calc.}}$

	Fluorides			Chlorides			Bromides			Iodides		
	$r_{\text{Obs.}}$	$r_{\text{Calc.}}$	Δr	$r_{\text{Obs.}}$	$r_{\text{Calc.}}$	Δr	$r_{\text{Obs.}}$	$r_{\text{Calc.}}$	Δr	$r_{\text{Obs.}}$	$r_{\text{Calc.}}$	Δr
Li	2·01	2·01	0·00	2·57	2·49	0·08	2·75	2·64	0·11	3·03	2·87	0·16
Na	2·31	2·31	0·00	2·81	2·79	0·02	2·98	2·94	0·04	3·23	3·17	0·06
K	2·66	2·66	0·00	3·14	3·14	0·00	3·29	3·29	0·00	3·53	3·52	0.01
Rb	2·82	2·81	0·01	3·29	3·29	0·00	3·43	3·44	−0·01	3·66	3·67	−0·01
Cs	3·00	3·00	0·00	3·56*	3·58	−0·02	3·71*	3·72	−0·01	3·95*	3·95	0·00

* Crystals having the CsCl type of structure (co-ordination number 8).

negative ions become important and should strictly be taken into account, as Pauling has attempted to do, in explaining the observed interatomic distances. It may be noted that these additional repulsions produce larger interatomic distances than the values calculated by adding the appropriate radii, and appear to be important in LiCl, LiBr, LiI, NaBr (?), and NaI.

In view of these considerations, we shall first examine the susceptibilities of crystals having the rock-salt type of structure in which the negative ions are not in specially close proximity, *i.e.*, we shall consider the following:—

LiF			
NaF	NaCl		
KF	KCl	KBr	KI
RbF	RbCl	RbBr	RbI
CsF			

It is not possible to obtain the susceptibilities of the individual ions directly from the experimental data; the problem is equivalent to trying to solve n simultaneous equations with $(n + 1)$ unknown quantities. Various ways have been suggested for introducing an additional equation, but they are all open to serious objections. It is not essential to obtain the ionic susceptibilities, however, in order to test their additivity. If the diamagnetic susceptibility of the smallest ion, Li$^+$, is taken as $x \times 10^{-6}$, the susceptibilities of the remaining ions can be found in terms of x; these are tabulated in Table IV and were obtained by attaching equal weight to the experimental results for the salts listed above.

TABLE IV—GRAM-IONIC SUSCEPTIBILITIES, $\chi \times 10^6$, IN TERMS OF THE SUSCEPTIBILITY OF LI$^+$

Li$^+$	x		
Na$^+$	$5{\cdot}4 + x$	F$^-$	$10{\cdot}1 - x$
K$^+$	$13{\cdot}9 + x$	Cl$^-$	$24{\cdot}9 - x$
Rb$^+$	$21{\cdot}3 + x$	Br$^-$	$35{\cdot}2 - x$
Cs$^+$	$34{\cdot}4 + x$	I$^-$	$51{\cdot}3 - x$

The results in Table IV can be compared with the experimental results on the assumption that the ionic susceptibilities are additive; the results are set out in Table V. A consideration of the data in this table shows that the susceptibilities are additive within the limits of experimental error in all the alkaline halides except LiCl, LiBr, LiI, and CsCl, CsBr, and CsI. In these exceptional cases, the observed susceptibilities are *less*

TABLE V—COMPARISON OF OBSERVED AND CALCULATED SUSCEPTIBILITIES, $\chi \times 10^6$

$\Delta\chi = \chi_{\text{Calc.}} - \chi_{\text{Obs.}}$

	Fluorides			Chlorides			Bromides			Iodides		
	Obs.	Calc.	$\Delta\chi$	Obs.	Calc.	$\Delta\chi$	Obs.	Calc.	$\Delta\chi$	Obs.	Calc.	$\Delta\chi$
Li	$10{\cdot}1_4$	10·1	0·0	23·3	24·9	1·6	34·0	35·2	1·2	50·0 (?)	51·3	1·3(?)
Na	15·6	15·5	—0·1	30·2	30·3	0·1	41·1	40·6	—0·5	57·0	56·7	—0·3
K	23·6	24·0	0·4	38·8	38·8	0·0	49·2	49·1	—0·1	65·7	65·2	—0·5
Rb	31·9	31·4	—0·5	46·4	46·2	—0·2	56·4	56·5	0·1	72·2	72·6	0·4
Cs	44·5	44·5	0·0	56·7	59·4	2·7	67·2	69·6	2·4	82·6	85·7	3·1

than those calculated, assuming additivity, and the differences are considerably greater than can be attributed to experimental errors.

5—A DISCUSSION OF THE DEPARTURE OF DIAMAGNETIC SUSCEPTIBILITIES FROM THE ADDITIVITY LAW

It is seen that the departures from additivity coincide (*a*) with a change of crystal structure, (*b*) with the near approach of the negative ions in crystals having small positive ions; in both cases the susceptibilities are lower than the values obtained by assuming additivity. It is of interest to consider first whether the lowering of the susceptibilities can be correlated with changes of the ionic packing sizes. In passing from the rock-salt type of structure to the caseium chloride type there is an *increase* in the ionic packing sizes. Also in LiCl, LiBr, and LiI, the additional repulsions between the negative ions make the effective packing sizes of the ions *greater*. Now in the case of free, spherically symmetrical ions, the gram-ionic susceptibility is given by the equation*

$$\chi = -\frac{Ne^2}{6mc^2}\Sigma\overline{r^2},$$

where $\overline{r^2}$ is the mean square distance of an electron from the nucleus and the summation extends over all the electrons; *e*, *m*, *c*, and N have their usual significance. If an increase in the packing sizes of ions led to an increase in $\Sigma\overline{r^2}$ for ions in crystals, then there should be an increase in susceptibility. The experimental results show that there is a decrease in susceptibility with an increase of packing size and therefore the two cannot be directly correlated.

The next question to consider is the possible effect of deformations of the ions arising from their close proximity in the crystal lattice. A comparison of the susceptibilities of ions in dilute solutions and in crystals shows that the values for crystals are about 2%–5% less than the values for the corresponding solutions. It is also well established† that the susceptibilities of ions in crystals are less than the susceptibilities of free ions calculated theoretically. It may be said, then, somewhat empirically, that the deforming actions which the ions have on each other when packed into a crystal lower their susceptibilities. An increase of co-ordination number by increasing the number of deformations per atom is therefore likely to decrease the susceptibility. This is what is found in CsCl, CsBr, and CsI. For CsI we have the following data: in aqueous solution, $\chi \times 10^6 = 92\cdot5$; for co-ordination number 6 (assuming additivity of the values in Table IV), $\chi \times 10^6 = 85\cdot7$; experimentally for co-ordination number 8, $\chi \times 10^6 = 82\cdot6$. The decrease in susceptibility for co-ordination number 6 is 6·8 and for co-ordination number 8 is 9·1; the decrease in susceptibility is therefore approximately proportional to the co-ordination number, which is what would be expected if the decrease is due to deformations arising from " contacts " between adjacent ions.

---

* *See*, for example, E. C. Stoner, "Magnetism and Matter," ch. iv, sect. 3, Methuen (1934).

† Stoner, *op. cit.*, ch. ix.

In the case of LiCl, LiBr, and LiI, the decrease of susceptibility cannot be correlated with an increase of co-ordination number but, owing to the special conditions existing in these crystals, there is effectively an increase in the number of " contacts " per atom. In LiCl, for example, there are surrounding each Cl^- ion, six Li^+ ions and also eight Cl^- ions which approach very closely; in other chlorides the positive ions are sufficiently large to keep the negative ions well spaced. In LiF the F^- ions, which are considerably smaller than Cl^- ions, are not brought into specially close proximity. It would appear, then, that the decreased susceptibility of LiCl, LiBr, and LiI arises from the deformations produced by the near approach of the negative ions.

6—DERIVATION OF IONIC SUSCEPTIBILITIES

As mentioned in the previous section, ionic susceptibilities cannot be derived directly from a series of results for binary compounds without the introduction of an additional assumption. A number of methods has been suggested for doing this, but they are all open to objection. Joos* first suggested dividing the susceptibility of a salt formed of similar ions, *e.g.*, CsI, in the ratio of the squares of the atomic numbers. This procedure lacks justification for the following reason: the gram-ionic susceptibility for spherically symmetrical atoms is given by

$$\chi = -(\mathrm{N}e^2/6mc^2)\,\Sigma\overline{r^2}.$$

For a single electron in the field of a nucleus of charge Ze, $\overline{r^2}$ is proportional to $1/Z^2$, but for an atom containing many electrons such is not the case, because of the screening of outer by inner electrons. This is very important, because it is the outer group of electrons which contributes mainly to the diamagnetic susceptibility. The use of effective nuclear charges with the correct weight attached to each electron shell, as was first suggested by one of us,* improves the method considerably; but uncertainties still remain, for, in the first place, it involves the use of an atomic model and, secondly, it involves the assumption that results calculated for free atoms and ions can be applied to ions in crystals. The latter assumption is specially doubtful because susceptibilities depend mainly on outer electrons, and these will be more affected than inner electrons when atoms are packed into crystals. Susceptibilities have been calculated by Pauling,† using hydrogen-like wave functions, by Stoner,‡ using Hartree's wave functions, and by Brindley,§ using Slater's wave functions. A full discussion of these results has been given recently by Stoner.|| The figures obtained from Slater's wave functions show the best agreement with experimental results for crystals; but this must be regarded as to some extent fortuitous for the reasons given above. Another method has been used by Weiss,¶ based on the value of the susceptibility of HCl in aqueous solution, but it involves a rather uncertain correction for the effect of the dissolved H^+ ions on the susceptibility of water; in any case the susceptibilities of solutions and crystals are not directly comparable. Angus** has recently calculated the susceptibilities of a large number of atoms and ions by the use of a modified form of Slater's method for obtaining approximate wave functions. The modification, however, is scarcely justifiable. The screening constants and effective quantum numbers given by Slater require that no distinction shall be made between *s* and *p* sub-groups with the same total quantum number; Angus separates these groups while retaining Slater's values. He also assumes, in effect, that *s* electrons in an *s*, *p* group are not screened by the *p* electrons, which cannot be strictly true. There is, of course, no objection to the separation of *s* and *p* electrons, but a modification of Slater's scheme would be required.

---

* ' Z. Physik,' vol. 19, p. 347 (1923); vol. 32, p. 835 (1925).

In view of these difficulties, we suggest the following method which appears to suffer from fewer objections than other methods. For the gram-ionic susceptibility of Li^+ in LiF the value $0 \cdot 7 \times 10^{-6}$ is assumed. This figure is arrived at from consideration of the following results: Pauling, using hydrogen-like wave functions, has given the value $0 \cdot 63 \times 10^{-6}$; Stoner, using Hartree's wave functions, has obtained

---

* Brindley, 'Phil. Mag.,' vol. 11, p. 786 (1931).

† 'Proc. Roy. Soc.,' A, vol. 114, p. 181 (1927).

‡ 'Proc. Leeds Phil. Soc.,' vol. 1, p. 484 (1929).

§ *Loc. cit.*

|| "Magnetism and Matter," ch. ix, Methuen (1934).

¶ 'J. Phys. Radium,' vol. 1, p. 185 (1930).

** 'Proc. Roy. Soc.,' A, vol. 136, p. 569 (1932).

$0 \cdot 70 \times 10^{-6}$, and Brindley, using Slater's wave functions, has obtained $0 \cdot 67 \times 10^{-6}$. Since the susceptibility of Li^+ is very small, quite a large percentage error in the value assumed will have little effect on the results for other ions. Using this value for Li^+ in conjunction with the data of Table IV, the values given in Table VI are obtained for the ionic susceptibilities.

TABLE VI—GRAM-IONIC SUSCEPTIBILITIES × 10^{-6} FOR IONS IN CRYSTALS OF THE ROCK-SALT TYPE

Li^+	0·7		
Na^+	6·1	F^-	9·4
K^+	14·6	Cl^-	24·2
Rb^+	22·0	Br^-	34·5
Cs^+	35·1	I^-	50·6

Finally, we wish to thank Professor F. H. Newman for his continued interest in this work, and Dr. E. C. Stoner, who has read the manuscript of the paper and made several useful suggestions.

7—CONCLUSIONS AND SUMMARY

This is the second of a series of papers on the diamagnetic susceptibilities of salts containing ions of rare gas type. This paper gives experimental results for the halides of lithium, rubidium, and caesium and discusses them in conjunction with results given in Part I for sodium and potassium halides. The experimental results are summarized in Table I; out of these 20 compounds, measurements for 10 of them have not previously been made. With these data we have been able to make a more thorough test of the additivity of diamagnetic susceptibilities than has hitherto been possible. It is shown that the susceptibilities are not additive throughout this series of salts and that the additivity law breaks down chiefly for LiCl, LiBr, and LiI, and CsCl, CsBr, and CsI. For the remaining salts the susceptibilities are additive within the limits of experimental error. The departures from additivity are considered in relation to similar results for ionic packing sizes; it is suggested that the low values found for the susceptibilities of the Li and Cs chlorides, bromides, and iodides are due to deformations of the ions produced (*a*) by the unusually close approach of the negative ions in LiCl, LiBr, and LiI, and (*b*) by the change of crystal structure in CsCl, CsBr, and CsI. A series of ionic susceptibilities is derived for ions in crystals of the rock-salt type in terms of $Li^+ = 0 \cdot 7 \times 10^{-6}$.

Investigations on the Combustion of Hydrocarbons

I—The Influence of Molecular Structure on Hydrocarbon Combustion

By A. R. UBBELOHDE, M.A. (Oxon), Department of Thermodynamics, Clarendon Laboratory, Oxford

(*Communicated by A. C. G. Egerton, F.R.S.—Received May* 24, 1935)

I—INTRODUCTORY

In the combustion of hydrocarbons the number of intermediary processes which might occur before a combustion, such as

$$C_5H_{12} + 8O_2 = 5CO_2 + 6H_2O,$$

is completed is very large. In order to collect some of the more important facts bearing on the subject from the large literature which is accumulating, a brief review is first made of the methods which have been adopted to attack the problem. The fundamental processes involved in the attack of a hydrocarbon molecule by oxygen will then be discussed. It then becomes possible to put forward a reasonable mechanism for the combustion of the longer chain hydrocarbons, further evidence for which is obtained in the experimental portion of the work set out in Part II.

(*a*) *Chemical*—The chemical method of studying combustion mechanism involves the isolation of intermediates, such as alcohols,[*][†] aldehydes,[‡] peroxides,[§][||] and acids. The formation of unsaturated substances and of hydrogen is likewise important. The principal aim of early theories of combustion, such as the hydroxylation theory,[¶] and to some extent the peroxide theory,[‡][**][††] is to account for the formation of these intermediates in terms of a combustion mechanism. For example, the oxidation of ethane according to the hydroxylation theory is written (*cf.* Bone)[¶]

---

[*] Landa, 'C.R. Acad. Sci., Paris,' vol. 186, p. 589 (1928).

[†] Newitt and Haffner, 'Proc. Roy. Soc.,' A, vol. 134, p. 591 (1932).

[‡] Callendar and Mardles, 'Engineering,' vol. 123, pp. 147, 182, 210 (1927).

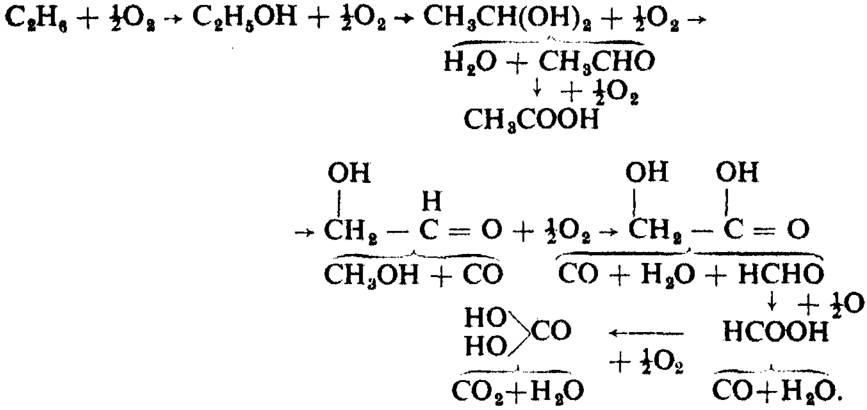
[§] Mondain Monval and Quanquin, 'Ann. Chim. (Phys.),' vol. 15, p. 309 (1931).

[||] Pease, 'J. Amer. Chem. Soc.,' vol. 56, p. 2034 (1934).

[¶] Bone, 'J. Chem. Soc.,' p. 1599 (1933).

[**] Egerton, 'J. Inst. Pet. Tech.,' vol. 13, p. 244 (1927); 'Nature,' vol. 121, p. 10 (1928).

[††] Moureu and Dufraisse, 'C.R. Acad. Sci., Paris,' vol. 185, p. 1545; vol. 186, p. 196 (1928).



$$
\begin{array}{l}
C_2H_6 + \tfrac{1}{2}O_2 \rightarrow C_2H_5OH + \tfrac{1}{2}O_2 \rightarrow CH_3CH(OH)_2 + \tfrac{1}{2}O_2 \rightarrow \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \overbrace{H_2O + CH_3CHO} \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \downarrow + \tfrac{1}{2}O_2 \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad CH_3COOH
\end{array}
$$

$$
\begin{array}{l}
\rightarrow \overbrace{\overset{\overset{OH}{|}}{CH_2} - \overset{H}{C} = O}^{} + \tfrac{1}{2}O_2 \rightarrow \overset{\overset{OH}{|}}{CH_2} - \overset{\overset{OH}{|}}{C} = O \\
\quad \overbrace{CH_3OH + CO} \qquad\quad \overbrace{CO + H_2O + HCHO} \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \downarrow + \tfrac{1}{2}O_2 \\
\qquad\quad {}^{HO}_{HO}\!\!>\!CO \quad \xleftarrow[+\tfrac{1}{2}O_2]{} \quad HCOOH \\
\qquad\quad \overbrace{CO_2 + H_2O} \qquad\qquad\quad \overbrace{CO + H_2O}.
\end{array}
$$

A number of facts suggest that the degration of a hydrocarbon to oxides of carbon and hydrogen is more complicated than this scheme would imply. In the first place, intermediates of great importance for combustion might be too unstable to isolate by chemical means ; their occurrence has been inferred by one of the methods described below. A second limitation to the scheme is that it involves either a series of ternary collisions at each step, $2C_2H_6 + O_2 = 2\,C_2H_5OH$, or the intervention of oxygen atoms. The temperature coefficient and rate of reaction of hydrocarbon combustions make a series of ternary collisions unlikely. The possible role of oxygen atoms is discussed later. A third limitation to this hydroxylation scheme is that it does not make clear why most hydrocarbon combustions are autocatalytic.[†] This autocatalysis means that intermediates promote the oxidation of the parent hydrocarbon, and are themselves reformed, so that the addition of important intermediates may affect both the induction period and the rapid combustion of hydrocarbons.

The fact that the chemical method can deal only with stabilized intermediates must be remembered in assessing the value of different theories of combustion. In a simple reaction, A → B, the concentration of active A molecules according to theories of reaction velocity[‡] does not usually exceed 1 in 10^{10} or more. In certain chain reactions, the concentration of chain carriers at any instant may be shown to be given by an expression, concen-

---

[*] Bone, '*J. Chem. Soc.*,' p. *1599 (1933).*

[†] Egerton and Gates, ' Rep. and Mem. Aero Res. Ctee.,' No. 1079 (1926).

[‡] *Cf.* Hinshelwood, " Kinetics of Chemical Change in Gaseous Systems," for examples.

tration of carriers $= \dfrac{\text{rate of reaction}}{\text{total number of collisions}}$ similar to that for simple reactions. Even in more complicated examples the concentration of chain carrier will usually not be detectable by ordinary chemical means unless this concentration rises abnormally by chain-branching leading to ignition.*

(*b*) *Reaction Kinetics*—Since most oxidations involving a rupture of a C-C linkage take place with increase of pressure, the static method of measuring rate of reaction has been used. The complex processes occurring with the higher hydrocarbons, with which this discussion is chiefly concerned, may, however, be completely missed if only the rate of pressure increase is studied.† Estimates of the influence of factors such as fuel/oxygen ratio, total pressure, area and nature of surface, etc., on the velocity of reaction can also be made from the critical reaction temperature.‡ In the flow method, the progress of reaction can be determined from the local temperature gradients due to the heat evolved by the reaction, and from the rate of consumption or production of O_2, CO, and CO_2 as determined by analysis.§

The reaction kinetics of nearly all hydrocarbon combustions show the following characteristic features:

(1) Practically all combustion processes are chain reactions, the chain length depending on a number of factors in a complicated way.

(2) Over a wide range of concentrations the velocity is much more affected by an increase in fuel concentration than by an increase in oxygen. The most simple explanation is that some sort of oxygenated complex is first formed, and must collide with a second fuel molecule within a finite time for the reaction to proceed.

(3) Increased surface area slows down the reaction (Pidgeon and Egerton, *loc. cit.*), indicating chain breaking at the walls ; there is also evidence, however, that the chains may be initiated at the walls.||¶

(4) Since the temperature coefficient actually becomes negative over certain regions of temperature, a complicated sequence of events is clearly involved.**†† The combustion of the paraffins in the low-temperature region from 250° to about 350° C shows a number of characteristic differences

---

* Semenov " Chemical Kinetics and Chain Reactions," Oxford, 1935.

† Rideal and Brunner, 'J. Chem. Soc.,' pp. 1162, 2824 (1928).

‡ Pidgeon and Egerton, 'J. Chem. Soc.,' p. 661 (1932).

§ Beatty and Edgar, 'J. Amer. Chem. Soc.,' vol. 56, p. 102 (1934).

|| Haber and Alyea, 'Z. phys. Chem.,' B, vol. 10, p. 193 (1930).

¶ Goldmann, 'Z. phys. Chem.,' B, vol. 5, p. 316 (1929).

** Beatty and Edgar, *loc. cit.*

†† Neumann and Aivazov, 'Nature,' vol. 135, p. 655 (1935).

from that at higher temperatures, which makes it convenient to discuss the low-temperature combustion separately. Even if the same steps are involved, their relative importance in the autocatalytic combustion is greatly altered at higher temperatures, giving the appearance of a different mechanism. In what follows attention is principally given to combustion in the low-temperature region, where complex oxidation intermediates still have reasonable stability.

(*c*) *Chemi-Luminescence and Emission Spectra*—Spectroscopic examination of the flash on ignition at high temperatures has shown that it is due to light emitted from excited radicals such as C — C, C — H, and especially — OH,* which are produced in large numbers at the momentary high temperature due to the liberation of the heat of reaction. The phenomena of chemi-luminescence referred to are observed at a much lower temperature, which accounts for the name of cool flames, and are particularly marked when the fuel is in excess.† With fuel air mixtures flowing through the furnace at six litres/hour, the phenomena observed have been summarized in Table I.

TABLE I

Gas	%	Luminescence °C.	Lower ignition °C.	Higher ignition °C.
CH_4	5·3	740	—	770 (CH_3OH at 540).
	18·5	690	—	730
C_2H_6	4·2	380–410	—	660
	8·9	350–370	—	675
C_3H_8	3·5	270–310	—	665
	6·3	265–305	—	672
	14·5	255·290	304–310	Not possible
C_5H_{12}	2·1	242–260	—	670
	4·05	221–240	262–292	681
	16·55	225–244	259–270	Not possible
C_6H_{14}	2·05	230–255	—	684
	5·20	220–244	255–270	672
C_7H_{16}	1·55	230–245	—	686
	5·25	210–232	259–269	695
C_8H_{18}	0·95	215–230	—	670
	2·10	200–220	252–270	684

In this table, due to Prettre, the temperatures in the second column indicate when luminescence is first visible and then becomes strong.

---

* Bonhoeffer and Haber, 'Z. phys. Chem.,' A, vol. 137, p. 263 (1928).

† *Cf.* reference in Prettre, 'Bull. Soc. chim. Fr.,' vol. 51, p. 1132 (1932).

The influence of substituents in the hydrocarbon molecule is noteworthy. A — CHO group enormously enhances the phenomenon, since CH_3CHO shows intense luminescence in the neighbourhood of 200°, and $CH_3 - CH_3$ none. An — OR group has a similar effect, as experiments with ethyl ether show. The introduction of an — OH group, a — COOH group, or a double bond lessens the tendency to luminescent combustion at low temperatures. It is not surprising that this influence is closely parallel to that on the knock-rating of fuels, since knock is due to the sensitizing by slow combustion of the last portion of the cylinder charge to burn.* Factors that favour low-temperature combustion before the flame has arrived also favour the production of pro-knocks.

The behaviour of homologues and isomers is significant. The tendency to low-temperature combustion and knock runs parallel with the largest length of unbranched chain of carbon atoms in the molecule. Thus the higher straight-chain homologues oxidize more easily, but, if further carbon atoms are added by branching, the molecule is more resistant to oxidation. The effect of impurities on chemi-luminescence is parallel to that on slow combustion in the low-temperature region ; metallic anti-knocks and benzene diminish it, and amyl nitrite, acetaldehyde, and di-ethyl peroxide enhance it.

In general, chemi-luminescence is due to the formation of active molecules, in this case in a chain process, and these molecules have time to emit their energy by radiation before it is dissipated as translational energy, or used in propagating the chain. Spectrographic studies† have shown that the luminescent flames of CH_3CHO, C_2H_5CHO, C_6H_{14} and $(C_2H_5)_2O$ all give the same bands, suggesting a source common to these different combustions. It is thus important to establish the chemical nature of the molecule responsible for chemi-luminescence. Since it must be common to these different reactions, it might be an excited oxygen molecule, excited CO or CO_2, or an excited — CO linkage forming part of molecules such as RCHO, ROOH, or radicals derived from them. The — C — H linkage seems excluded by the nature of the ultra-violet absorption of the hydrocarbons. Inspection of published band systems shows that although the Runge bands for oxygen‡ show certain coincidences, the position and general distribution of intensities of the chemi-luminescence bands (from 3385–4903 A) correspond closely with the fluorescence spectrum of formaldehyde,* though there is an average shift of 25 A to the red in cold flames. This production of activated formaldehyde is thus a fairly general feature of low-temperature combustion, and is discussed in § 3. (I am indebted to Professor A. Fowler, F.R.S., and Mr. R. Pearse for the comparison of Emelèus's spectra with those of Gradstein.)

---

* Egerton, Llewellyn Smith, and Ubbelohde, 'Phil. Trans.,' A, vol. 234, p. 433 (1935).

† Emeléus, 'J. Chem. Soc.,' p. 2948 (1926) ; p. 1733 (1929).

‡ Lochte, Holtgreven and Dicke, 'Ann. Physik,' vol. 3, p. 937 (1929).

(*d*) *Induction Period and Wall Effects*—These two effects are so characteristic that they must be included in any discussion of the mechanism of hydrocarbon combustion. When a combustible mixture is introduced into a " clean " reaction vessel at a suitable temperature, reaction does not always start as soon as the gas has reached the temperature of the enclosure. If the progress of the reaction be measured by the change in pressure, no change may be observed for half an hour or more, and, with ethane, analysis shows† that no appreciable combustion has taken place, though it is clear that something must be happening on a very small scale.‡§ Once the reaction starts it proceeds smoothly.

The existence of an induction period suggests that the mixture must be sensitized in some way for combustion to proceed. This is confirmed by the fact that the addition of 2% of H_2O, I_2, C_2H_5OH, HCHO, CH_3CHO, and NO_2 greatly reduces the induction period,†‖ whereas mild anti-knocks such as aniline prolong it. Starting with a pure hydrocarbon/air mixture, the sensitizers which are presumably formed during the induction period may either accumulate in the gas phase, or may be needed to bring the walls of the vessel into a more suitable state for initiating chains. The relatively small amounts of sensitizers required suggest this second possibility, especially as chains are known to start from the walls in some cases (*see* above, and *cf.* Hadman, Thompson, and Hinshelwood).¶ A separation of the two effects, however, is possible in gases flowing through a heated tube. After a sufficient lapse of time, the walls are in a steady state with respect to adsorbed water, aldehydes, etc., but the same interval is still required for the autocatalytic accumulation in the gas phase, so that rapid reaction only sets in some way down the tube, the distance from the point of entry varying with the velocity of flow. A general survey of flow methods leaves no doubt that the induction phenomena involve not only sensitization of the walls but also accumulation of sensitizers in the gas phase.

---

* Gradstein, ' Z. phys. Chem.,' B, vol. 22, p. 384 (1933).

† Bone and Hill, ' Proc. Roy. Soc.,' A, vol. 129, p. 434 (1930).

‡ Egerton, ' Trans. Faraday Soc.,' vol. 28, p. 412 (1932).

§ Semenov, " Chemical Kinetics and Chain Reactions," Oxford (1935).

‖ Rideal and Brunner, ' J. Chem. Soc.,' pp. 1162, 2824 (1928).

¶ ' Proc. Roy. Soc.,' A, vol. 137, p. 97 ; vol. 138, p. 297 (1932).

The problem is made more difficult by the fact that the walls undoubtedly slow down combustion by breaking reaction chains. A number of ways of controlling the state of the walls has been reported, *e.g.*, (*a*) repetition of a combustion in the same apparatus gives a reproducible induction period, in place of erratic results in a " clean " vessel ; (*b*) washing with suitable chemicals affects both the combustion, and the products, which can be isolated. Thus, washing a quartz surface with HF lowers the temperature of reaction for ethane by 150°,* and sintering has the opposite effect. Washing with potassium chloride slows down the reaction, but lessens still more the amount of H_2O_2 or organic peroxides which can be isolated.† Washing with dilute nitric acid has the opposite effect.

(*e*) *Positive and Negative Catalysts*—The principal negative catalysts for low-temperature combustion are the well-known anti-knocks ; both derivatives of metals with two stages of oxidation, and easily oxidized amines, etc., slow down the reaction. The small amount which is effective (1 part in 1000 of fuel) is significant evidence for their stopping processes, which involve either long chains or at least require a minimum time interval before branching occurs (*see* later). The chemical nature of the anti-knocks suggests that they destroy an oxygen-rich carrier and prevent the oxidation of further fuel molecules. The kinetic evidence for such an oxygen-rich carrier has already been mentioned. At high temperatures and in the flame these metallic inhibitors have little or no action. This may possibly be due to a different mechanism of combustion at these temperatures, or to the fact that the equilibrium between two stages of oxidation of the metal is displaced too far in one direction‡; both effects may, of course, operate simultaneously.

The principal positive catalyst is nitrogen peroxide. This accelerates the combustion of a number of compounds, such as H_2, CO, CH_4, C_2H_6, C_2H_2, alcohols, and benzene, so that its action is very general. Oxidation catalysed by nitrogen peroxide may be due to the fact that oxides of nitrogen can change their valency easily by adding H or losing O (*see* below). Aldehydes also accelerate hydrocarbon combustion, and since they are one of the first detectable products, the autocatalytic nature of combustion is partly due to their formation. This effect of aldehydes is important in suggesting a similar carrier in the chain mechanism of hydrocarbon and aldehyde combustion, while, owing to the lower temperature and photochemical properties of aldehyde oxidation, information on its chain mechanism is more readily available.*

---

* Kowalsky, 'Phys. Z. Sowjet,' vol. 1, pp. 451, 567 (1932).

† Pease, 'J. Amer. Chem. Soc.,' vol. 56, p. 2034 (1934).

‡ Egerton, 'J. Inst. Pet. Tech.,' vol. 13, p. 244 (1927) ; 'Nature,' vol. 121, p. 10 (1928).

To sum up this discussion, the principal facts any theory of combustion has to explain are the initiation of the process, the nature of the chain carriers, and what intermediates of oxidation are likely to be stabilized. A second important group of questions arises from a study of the oxidation of homologues and isomers. The profound effect of certain substituents, such as — CHO and — OR, on the ease of oxidation of hydrocarbons has already been mentioned. A kindred question is the point of attack of the molecule. For aliphatic hydrocarbons attack at the end methyl group seems to occur most readily,† and with substituted benzenes the point of attack is the α carbon atom.‡ Similar remarks probably apply to the oxidation of ethers.§|| Energy considerations may be applied to all these questions.

II—ENERGY FACTORS IN THE COMBUSTION OF HYDROCARBONS

According to Slater,¶ the energy states of a molecule may be written in the form $E = Q + \Sigma a_p - \frac{1}{2} \Sigma' a_{11}$, where the first term includes forces due to electric charges, permanent dipoles, and van der Waals effects, the second-term resonance attractions between pairs of electrons forming covalent bonds, and the third-term resonance between unbound atoms. The first and third terms thus include the effect of neighbouring atoms on the chemical linkages characterized by the second term. When these are of the order of a few thousand kg cal, the factor $e^{-E/RT}$ over the range of temperatures 0° — 300° C will range from 10^{-1} to 10^{-3}.

If we consider a series of homologues undergoing the same primary process, *e.g.*, $RCH_2Cl + H_2O = RCH_2OH + HCl$, the effect of substituents in R on the primary activation at the — C — Cl linkage will usually not exceed a few thousand calories, leading to relative reaction rates in homologues and substitution products differing at the most by 100 : 1. The vast material on organic reactivity is mainly concerned with such relative reaction rates. Provided the primary process is unchanged, these give a very fair measure of the influence of distant groups on a given linkage, without requiring a detailed analysis of the kinetics. The repulsion effects due to neighbouring atoms, and dipole moments of substituents will affect

---

* Bäckstrom, 'Z. phys. Chem.,' B, vol. 25, p. 99 (1934).

† Pope, Dykstra, and Edgar, 'J. Amer. Chem. Soc.,' vol. 51, pp. 1875, 2203 (1929).

‡ Stephens, 'J. Amer. Chem. Soc.,' vol. 48, p. 1824 (1926).

§ Wieland and Wingler, 'Liebig's Ann.,' vol. 431, p. 317 (1923).

|| Milas, 'J. Amer. Chem. Soc.,' vol. 53, p. 221 (1931).

¶ *Cf.* Eyring, 'J. Amer. Chem. Soc.,' vol. 54, p. 3191 (1934).

the first and third terms of the expression for the energy of the molecule, whereas tautomeric effects raise or lower the energy of activation by affecting the second term.

In combustion reactions, the marked differences in the series hydrocarbon, alcohol, ether, is probably due to the O atom. In the alcohols its effect is probably directed towards the H of the hydroxyl, either as a dipole or tautomeric effect, whereas in ethers it must necessarily affect the carbon atom, and thus facilitate the oxidation of the α C — H linkage. Since substituted benzenes show marked preferential oxidation in the α position, it may be that a tautomeric effect is involved in the attack by oxygen.

The ease of combustion of aldehydes might be partly ascribed to similar influences, but is mainly due to the ease of formation of free radicals (*see* below). Ketones are not oxidized as easily as aldehydes.

In the long chain paraffins dipole and tautomeric effects can hardly explain differences in ease of combustion. Furthermore, a relative loosening of specific — C — H linkages by repulsion effects (*cf.* Eyring, *loc. cit.*) is greatest for secondary and tertiary C atoms, which nevertheless show the least tendency to oxidation, so that if steric hindrance is important, the branched isomers should show larger temperature coefficients of combustion. (The fact that they have lower boiling-points indicates considerable shielding of the van der Waals attractions of the innermost atoms by the outer atoms in a branched isomer.)

However, such steric hindrance cannot explain why the higher paraffins are much more easily oxidized than methane and ethane ; furthermore, the principal attack at the end of a long chain paraffin, and the fact that more aldehydes are formed than ketones, suggest that another factor of considerable importance must be included in explaining the reactivity of hydrocarbons. The knock rating of the naphthenes is higher than that of corresponding paraffins, and if a double bond be introduced in the middle of a long chain the knock rating rises to that of the longest aliphatic chain left.* A rise in rating means less tendency to knock.

In order to explain facts of this kind, it is necessary also to consider the vibrational energy content of the hydrocarbon. This may be computed in terms of the natural frequencies of the molecule. Starting with methane and ethane, and gradually interposing more CH_2 groups, the molecule will behave as a coupled oscillator, and will thus give rise to frequencies lower than, *e.g.*, the fundamental C — C vibration. In consequence, the excitation of vibrational energy below 350° C is proportionately much larger for the longer molecules, whereas increasing the number of carbon atoms by branching would not have this effect. Furthermore, the coupling would be interrupted at the double bond. One mode of " action at a distance " in hydrocarbon molecules depends on the possibility of vibrational coupling ; neglecting other modes of distant action already referred to, the details of vibrational coupling may be developed for the paraffins, where a number of experimental results suggest that it is an important factor in their combustion.

---

* Lovell, Campbell, and Boyd, ' J. Ind. Eng. Chem.,' vol. 25, pp. 555, 1106 (1933).

One consequence will be that the total null-point energy per C – C linkage is smaller for the long molecules. This lowers the apparent energy of addition of successive CH_2 groups, though the effect rapidly tends to zero for the longer molecules.

Vibrational Model of the Hydrocarbon Molecule—Adequate models of the hydrocarbon molecule cannot yet be constructed, but the principal results may be obtained from quite simple considerations.* In a molecule $(CH_2)_{2n+2}$ with all the valency angles 180°, in which the vibrations are due to the elongation and compression of the valency bonds, the vibrations are assumed to be simple harmonic, and the restoring force on any group is taken as proportional to its displacement relative to its two immediate neighbours only. For the present purpose, the CH_2 groups are assumed to move as a whole in changes of length of the C – C bonds.

If m is the mass of the CH_2, $2\pi k$ the force constant for the stretching of a C – C linkage, X_r the displacement of the rth radical from its equilibrium position, then for a normal frequency ω

$$\omega^2 X_r = k/m\,[(X_r - X_{r-1}) - (X_{r+1} - X_r)], \tag{1}$$

and for the two end radicals

$$\omega^2 X_1 = -\,k/m\,(X_2 - X_1), \tag{2}$$

and

$$\omega^2 X_{2n+2} = k/m\,(X_{2n+2} - X_{2n+1}). \tag{3}$$

Choosing as variable the actual elongation y_s of the link between the rth and $r + 1$th atoms (s being the number of the link), and renumbering the $2n + 1$ linkages between the $2n + 2$ atoms (n may be a whole or half-number) from $- n$ to $+ n$, we obtain by subtraction

$$\omega^2 Y_s = \frac{2k}{m} Y_s - k/m\,Y_{s-1} - \frac{k}{m} Y_{s+1} \tag{4}$$

$$\omega^2 Y_n = \frac{2k}{m} Y_n - \frac{k}{m} Y_{n-1} \tag{5}$$

$$\omega^2 Y_{-n} = \frac{2k}{m} Y_{-n} - \frac{k}{m} Y_{-n+1}. \tag{6}$$

---

* Bartolomé and Teller, ' Z. phys. Chem.,' B, vol. 19, p. 366 (1932).

The solution to equation (4) is $Y_s = \sin\frac{2\pi s}{\lambda}$ or $\cos\frac{2\pi s}{\lambda}$. Substituting in (4), and dividing by $\sin\frac{2\pi s}{\lambda}$ after rearranging,

$$\omega^2 = \frac{2k}{m}\left(1 - \cos\frac{2\pi}{\lambda}\right), \qquad i.e., \ \omega = 2\sqrt{\frac{k}{m}}\sin\frac{\pi}{\lambda}.$$

To satisfy (5) and (6)

$$Y_{n+1} = 0 = Y_{-(n+1)}$$

so that

$$\sin\frac{2\pi(n+1)}{\lambda} = 0 = \sin j\pi, \text{ where } j \text{ is a whole number,}$$

and

$$\frac{2\pi}{\lambda} = \frac{\pi j}{(n+1)} \qquad \begin{array}{l} n \geqslant j > 0 \text{ if } n \text{ is whole.} \\ \frac{2n+1}{2} \geqslant j > 0 \text{ if } n \text{ is a half-number.} \end{array}$$

The other solutions are

$$\cos\frac{2\pi(n+1)}{\lambda} = 0$$

or

$$\frac{2\pi}{\lambda} = \frac{(2j+1)\pi}{2(n+1)} \qquad \begin{array}{l} n \geqslant j \geqslant 0 \text{ if } n \text{ is whole.} \\ \frac{2n+1}{2} \geqslant j \geqslant 0 \text{ if } n \text{ is a half-number.} \end{array}$$

In all, this gives $2n + 1$ solutions, which is equal to the number of degrees of freedom of vibration for a linear molecule. The normal frequencies are thus given by the general equation

$$\omega = 2\sqrt{\frac{k}{m}}\sin\frac{(2j+1)}{4(n+1)}\pi \qquad \begin{array}{l} 2n+1 \geqslant j \geqslant 0 \text{ if } n \text{ is a half-number,} \\ 2n \geqslant j \geqslant 0 \text{ if } n \text{ is a whole number,} \end{array}$$

where $\omega = \sqrt{\frac{2k}{m}}$ is the frequency of vibration of a single link.

Heat of Formation of Chain Molecules—The heat of formation is determined from a combustion process—

$$C_nH_{2n+2} + (3n+1)/2O_2 = n\,CO_2 + (n+1)\,H_2O \ \ldots\ldots \qquad + H_0$$

which may be decomposed into the processes,

$$C_nH_{2n+2} = 2\,CH_3 + (n-2)\,CH_2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad + H_1$$
$$2\,CH_3 + 7/2\,O_2 = 2CO_2 + 3H_2O \ \ldots\ldots\ldots\ldots\ldots\ldots \qquad + H_2$$
$$(n-2)\,CH_2 + 3/2\,(n-2)\,O_2 = (n-2)\,CO_2 + (n-2)\,H_2O \qquad + H_3$$

where $H_0 = H_1 + H_2 + H_3$.

In the above equations H_2 and H_3 are linear functions of the number of CH_2 groups. Even if we neglect repulsion effects of the type described by Eyring (*loc. cit.*), the heat, H_1, will not be a linear function of the number of CH_2 groups, owing to the fact that the vibrational energy content of the hydrocarbon has to be expressed by the equation

$$H_{\text{vib}} = \Sigma \frac{Nh\nu}{e^{h\nu/kT} - 1} + \frac{Nh\nu}{2} \quad \text{per mole.}$$

The first term sums the Planck-Einstein functions for the natural frequencies of the molecule, and is usually approximately corrected for when H_0 is reduced to the value at 0°K, by allowing for the specific heats.* The second term refers to the null-point energy of the hydrocarbon chain, and has to be subtracted from the binding energy of the CH_2 groups. The result is that when a coupling of groups permits natural frequencies lower than the primary C — C or C — H vibration, the addition of more CH_2 groups apparently leads to firmer binding.

For the chain molecule already discussed, the null-point energy per mole due to C — C linkages will be

$$E_0 = Nh\sqrt{\frac{k}{m}} \sum_0^{2n} \sin \frac{2j+1}{4(n+1)}\pi$$

which sums to

$$E_0 = Nh\sqrt{\frac{k}{m}} \sin\frac{\pi}{4} \cdot \sin\frac{(2n+1)\pi}{8(n+1)} \Big/ \sin\frac{\pi}{8(n+1)}.$$

If we estimate the binding energy C — C from that observed from the addition of CH_2 groups when $n \gg 1$, the null-point energy per linkage will appear to be constant and has the value

$$E_0/(2n+1) = \operatorname*{Lt}_{n \gg 1} Nh\sqrt{\frac{k}{m}} \sin\frac{\pi}{4} \cdot \sin\frac{(2n+1)\pi}{8(n+1)} \Big/ (2n+1)\sin\frac{\pi}{8(n+1)}$$
$$= \frac{2Nh}{\pi}\sqrt{\frac{k}{m}},$$

whereas the null-point energy for a single pair of C atoms is

$$\frac{2Nh}{2\sqrt{2}}\sqrt{\frac{k}{m}}.$$

The drop per link is thus 10%, and practically reaches the average value for $n \gg 1$ for about five or more links.

---

* Rossini, 'Bur. Stand. J. Res.,' vol. 12, p. 735 ; vol. 13, pp. 21, 189 (1934).

Although the linear molecule vibrates in one dimension only, the actual hydrocarbon will show a decrease in average null-point energy per link which includes every mode of vibration for which coupling takes place on increasing the number of carbon atoms. Accurate estimates of the influence of null-point energy on heats of formation can only be made from a knowledge of the actual frequencies, when these have been sufficiently elucidated from Raman and infra-red measurements* ; an appreciable influence will only be observed when coupling between vibrations of high frequency can take place. In the data on the heat of formation of the alcohols and hydrocarbons (Rossini, *loc. cit.*), the alcohols show an increased binding energy per CH_2 group, tending asymptotically to constant binding energy. This is what would be expected from the effect just discussed. The normal hydrocarbons, however, show an inverted behaviour, and in this case effects, such as repulsion, particularly in replacing the first H atom in methane, must preponderate over the lowering in average null-point energy. Even in this case, however, the altered symmetry of the CH_3X might force different vibrations on the three remaining hydrogen atoms, leading to a null-point energy correction in addition to repulsion effects.

Branched Isomers and Cyclic Paraffins—Even if the fundamental combustion reaction involving the C — H link is the same in isomers, the activation energy available will depend on the vibrational content of the whole molecule, which is largest in the longest chains at ordinary temperatures. The simplest schematic model for discussing the effect of branching and ring closure on ease of combustion is a zigzag chain with fixed angle β between the links. The forces on the neighbouring atoms due to compression or elongation will be lowered in the ratio $\cos\beta : 1$. Equations (4), (5), (6) now become

$$\omega^2 Y_s = \frac{2k}{m} Y_s - \frac{k}{m} \cos\beta \, (Y_{s-1} + Y_{s+1}) \tag{4'}$$

$$\omega^2 Y_n = \frac{2k}{m} Y_n - \frac{k}{m} \cos\beta \, Y_{n-1} \tag{5'}$$

$$\omega^2 Y_{-n} = \frac{2k}{m} Y_{-n} - \frac{k}{m} \cos\beta \, Y_{-n+1}, \tag{6'}$$

and we have on substituting

$$Y_s = \sin\frac{2\pi s}{\lambda} \text{ or } \cos\frac{2\pi s}{\lambda}$$

$$\omega^2 = \frac{2k}{m}\left(1 - \cos\beta\cos\frac{2\pi}{\lambda}\right)$$

---

* Kohlrausch and Koppel, 'Z. phys. Chem.,' B, vol. 26, p. 209 (1934).

for the tetrahedral angle

$$\cos \beta = \tfrac{1}{3}$$

and

$$\omega^2 = \frac{2k}{m}\left(1 - \tfrac{1}{3} \cos \frac{2\pi}{\lambda}\right).$$

In general, for the simple chain the conditions are as before, though the numerical values of the solution are altered. For a cyclic paraffin, β is the angle in the ring, and the equations are similar, save that (numbering the links consecutively from 1 to $2n + 1$) $Y_{2n+1} = Y_1$. For such a closed chain,

$$\sin \frac{2(2n+1)\pi}{\lambda} = \sin \frac{2\pi}{\lambda}$$

and

$$\cos \frac{2(2n+1)\pi}{\lambda} = \cos \frac{2\pi}{\lambda}.$$

From these equations, the vibrational energy content of the naphthene will not be very different from the corresponding paraffin ; the smaller reactivity is in part due to a cause which is discussed below.

For branched molecules, similar equations,

$$\omega^2 Y_s = \frac{2k}{m} Y_s - \frac{k}{m} \cos \beta \, (Y_{s-1} + Y_{s+1})$$

$$\omega^2 Y_n = \frac{2k}{m} Y_n - \frac{k}{m} \cos \beta \, Y_{n-1}, \text{ etc.,}$$

hold for any part of the molecule not involving a point of branching. At the point of branching we must have

$$\omega^2 Y_s = \frac{2k}{m} Y_s - \frac{k}{m} \cos \beta \, (Y_{s+1} + Y_{s-1} + Y'_{s+1})$$

and two similar equations for the other two branches. Taking as solution

$$Y_s = \sin \frac{2\pi s}{\lambda} \text{ or } \cos \frac{2\pi s}{\lambda},$$

the result is the same as for unbranched chains if the point of branching is a node, *i.e.*, numbering the penultimate atoms of the three chains at a branch

$$p, q, r, \; Y_{p+1} = Y'_{p+1} = Y_{q+1}, \text{ etc.} = 0.$$

This gives three independent sets of vibrations in the three branches, the maximum wave-length corresponding with the greatest unbroken chain. The total number of solutions is the same as before. A double bond will in general have its force constant too different from that of C — C to permit coupling of the C — C vibrations across it (*cf.* the effect of substituents discussed by Bartholomé and Teller, *loc. cit.*), and will lead to a reactivity and generally a knock-rating of the aliphatic parts corresponding to their separate lengths.

From what has already been stated, the effect of adding CH$_2$ groups to methane and ethane is similar, in facilitating combustion, to that of introducing substituents such as — OR, phenyl, and — CHO. A characteristic difference is, however, that these substituents appear to increase the oxidizability of the C — H immediately next to them, as must occur when the lowering of the activation energy of the combustion process is due to tautomeric or electrical effects. With long-chain hydrocarbons, on the other hand, the greater ease of oxidation is due to a greater vibrational energy content, and oxidation would take place chiefly at the end methyl group, as the following considerations make clear.

The effect of increasing the vibrational energy of the molecule affects the reaction velocity of some fundamental process in hydrocarbon combustion in two ways. In the first place, the statistical probability of sufficient activation energy *e* being found at *any* C — C or C — H linkage is a function of the total energy. Following a theorem of Lorentz,* this probability is $\omega = \left(1 - \frac{e}{E}\right)^{3n-1}$ where E is the total energy, and *e* the energy required at any one bond, there being *n* bonds in all. At the temperatures where slow combustion is observed in the low temperature region, the fundamental vibration of the C — C bond is not fully excited, since the characteristic temperature for valency vibrations ($\nu \triangleq 900\ \text{cm}^{-1}$), for example, is 1030° C. E increases rapidly with the first few carbon atoms in a straight chain or ring, since, as shown, longer wave-lengths are then possible, whereas increasing the number of atoms by branching of the molecule does not lead to this effect. The exponent in the expression for ω likewise increases with the number of carbon atoms.

In addition to the general probability of finding sufficient activation energy at any bond in the molecule, end groups are in a particularly favourable position for converting translational and vibrational energy of the whole molecule into potential energy of stretching a given bond during actual collision, *e.g.*, with an oxygen molecule. In a chemical reaction involving adiabatic activation, a collision may be defined as the period during which kinetic energy of translation and vibration is converted into potential energy of position. This interval will in general be of the same order as the vibrational period of the bond unless the repulsion forces between, for example, O_2 and C — H are very different from those governing the oscillation of the — C — H itself. In consequence, if the hydrocarbon molecule is struck at one end in a suitable direction, sufficient time is available for all the natural vibrations to lead to a stretching of the bond, whereas for a collision with a central — CH_2 group the motions of the atoms partly cancel out. (An analogy is an organ pipe where only the open ends are positions of antinodes for both fundamental and all overtones.) For cyclic molecules, all positions are equally unfavourable, which may in part explain why these show a greater resistance to oxidation than the corresponding paraffin, although their vibrational energy content would seem to be not very different.

---

* Polanyi and Wigner, 'Z. phys. Chem.,' A, vol. 139, p. 439 (1928).

Since the combustion of hydrocarbons involves a stepwise attack, the products will, as a rule, have natural frequencies not very different from those of the parent molecule, particularly when the number of C atoms exceeds four or five. Optimum conditions are thus present for the return of any vibrational energy of the reaction products to the parent hydrocarbon, provided this process can take place in a time small compared with the time of loss due to other causes, *e.g.*, as translational energy. This emphasizes the necessity for having a large concentration of hydrocarbon molecules for combustion to proceed rapidly, since this favours the conservation of vibrational energy.

Similar arguments apply to any form of energy with several modes of dissipation, only one of which favours the continuation of the reaction chain. In such reactions a critical effect of pressure on reaction velocity will be observed when the frequency of collisions effective in propagating the chain becomes of the same order as the average life of the excited product. A typical case is the ignition of a thoroughly dry carbon monoxide oxygen mixture, which only proceeds when the pressure is sufficiently high. Here there is definite evidence of an excited product which must hand on its energy within a finite time for the chain to proceed.* It is suggested that the critical influence of pressure on the ignition of the higher hydrocarbons may be due to the intervention of an active product of short life in the low-temperature combustion, which can lead to chain-branching. In the experiments described by Townend and co-workers,† a critical effect of pressure is observed in the region 1–2 atmospheres. Since the distance diffused by a chain carrier rapidly decreases with rise in pressure,‡ it is unlikely that ignition suddenly becomes possible owing to less chain-breaking at the walls, as is the case, for example, for $H_2 + O_2$ at a few mm

* Ubbelohde, 'J. Chem. Soc.,' p. 972 (1933).

† Townend, Coehn, and Mandelkar, 'Proc. Roy. Soc.,' A, vol. 146, p. 113 (1934).

‡ Ubbelohde, 'Nature,' vol. 131, p. 328 (1933).

pressure. Furthermore, dilution with oxygen has an unfavourable effect, though it should act like an inert gas if chain-breaking at the walls is a deciding factor. The wall effects which have been noted are probably connected with the initiation of chains, rather than their interruption. The effect of anti-knocks is precisely what would be expected if these destroy the active product of short life which decides the course of combustion at low temperatures.

In conclusion, it must be emphasized that although the behaviour of hydrocarbons in the low-temperature region of combustion fits in very well with what might be expected if a minimum vibrational energy content is required, other factors undoubtedly enter ; some of these are discussed in the next section.

3—CHAIN MECHANISM IN THE LOW-TEMPERATURE COMBUSTION OF HYDROCARBONS

Several facts are known about hydrocarbon combustion which throw some light on the mechanism of the process.

(1) The combustion of a pure paraffin hydrocarbon is difficult to initiate, and probably starts heterogeneously unless, say, an aldehyde is added.

(2) A close similarity exists between the mechanisms of aldehyde and hydrocarbon combustion, suggesting a similar chain process.

(3) The addition of quite small amounts of negative catalysts (1 in 10^3 of fuel) slows down combustion markedly in both cases.

(4) Valuable evidence is also available from the phenomenon of knock in an internal combustion engine.

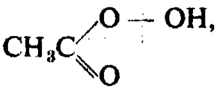
Since " knock " is due to sensitization of part of the charge by slow combustion (p. 358), before the flame has time to arrive, the addition of negative catalysts suppresses knock. An important difference from ordinary slow combustion lies in the fact that the last portion of the charge to burn is exposed to high temperatures and pressures for a relatively short time, and is also bombarded with active radicals, etc., from the flame front.

As a result, the addition of a number of pro-knocks in correspondingly small amounts (1 in 10^3 of fuel) has a marked effect on knock, but much less on slow combustion. This is probably due to the fact that when collision with a sufficiently energetic molecule occurs, pro-knocks disrupt to give free radicals which actively promote combustion. So far as has been ascertained, the only pro-knocks effective in these small concentrations, all have the possibility of a common mode of disruption, for which there is independent evidence in a number of cases,* and which is analogous to the process of chain-branching in the explosive combustion of hydrogen and deuterium:

---

* Ubbelohde and Egerton, ' Nature,' vol. 135, p. 67 (1935).

$R - O \vdots OR$, $R - O \vdots OH$,

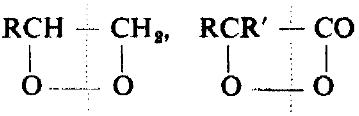


$RO \vdots NO$,

$RO \vdots NO_2$, $HO \vdots OH$, $DO \vdots OD$.

Chemically related peroxides and nitro-compounds which are unable to undergo such a disruption do not act as pro-knocks, which may explain why olefines and ketens, though they peroxidize much more easily, do not knock:

NO_2, $C_6H_5 - NO_2$, $CH_3 - NO_2$,

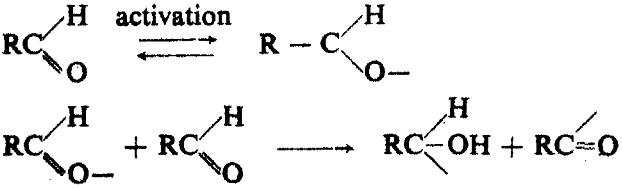


$$\begin{array}{ccc} RCH & \vdots & CH_2, \\ | & & | \\ O & \vdots & O \end{array} \quad \begin{array}{ccc} RCR' & \vdots & CO \\ | & & | \\ O & \vdots & O \end{array}$$

These pro-knocks would have a similar effect on low-temperature combustion, were it not for the fact that they all contain groups which greatly activate their own oxidation. Thus diethyl peroxide ($C_2H_5-O-O-C_2H_5$) is even more inflammable than ethyl ether at low temperatures, as would be expected from what has been said above. Pro-knocks will not disrupt till they collide with sufficiently energetic molecules, and in slow combustion the greater proportion is burnt away before this can take place.

The removal of pro-knocks by their own oxidation probably explains the fact that the pro-knock effect of diethyl peroxide in concentration 10^{-5} molfraction is inhibited by a concentration 10^{-5} of lead tetraethyl. The number of direct collisions between the two molecules is from one to twenty a second at the concentrations arising in the cycle of an internal combustion engine, and since the whole cycle is complete in, say, 1/10th sec, and the slow combustion occupies only some 3×10^{-8} sec, the inhibiting effect can hardly be due to direct collisions between anti- and pro-knock. The presence of the lead tetraethyl does, however, prevent the formation of energy-rich molecules by stopping the slow-combustion chain at an early stage, so that pro-knocks never get disrupted in time to share in the main combustion process.

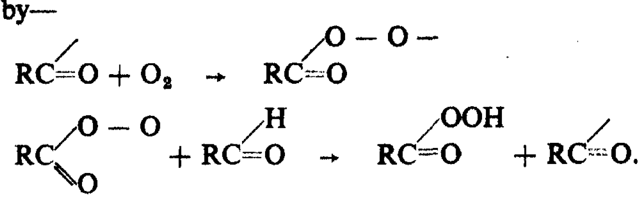
A chain mechanism may be proposed for discussion in view of these facts. According to Bäckstrom,* the aldehyde oxidation mechanism is—



---

* 'Z. phys. Chem.,' B, vol. 25, p. 99 (1934).

followed by—



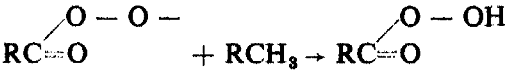
$$R\overset{/}{C}{=}O + O_2 \rightarrow R\overset{/O-O-}{C}{=}O$$

$$R\overset{/O-O}{\underset{\backslash O}{C}} + R\overset{/H}{C}{=}O \rightarrow R\overset{/OOH}{C}{=}O + R\overset{/}{C}{=}O.$$

This chain already shows the characteristic of most combustions that, unless the primary activation is followed by collision with a second fuel molecule within a finite time, reversion to a stable form takes place. When a suitable collision does occur, a chain carrier $R\overset{/}{C} = O$ of relatively long life and easier attack by oxygen is formed.

With hydrocarbons this may be extended to include the numerous additional possibilities, *e.g.* (stabilized products are underlined),

Initiation



$$R\overset{/O-O-}{C}{=}O + RCH_3 \rightarrow R\overset{/O-OH}{C}{=}O + RCH_2 + 10 \text{ kg cal} \quad (1)$$

(side reactions)

$RCH_2 + X$ (third body) $\rightarrow X + CH_3 + \underline{R''CH = CH_2}$ $- 17$ kg cal (1A)*

$RCH_2 + X \rightarrow X + H + \underline{R'CH = CH_2}$ $- 40$ kg cal (1A)*

Propagation

$RCH_2 + O_2 \rightarrow RCH_2 - O - O -$ $+ 13$ kg cal (2)

$RCH_2O - O - + RCH_3 \rightarrow \underline{RCH_2OOH} + RCH_2$ $+ 10$ kg cal (3)

The change $RCH_2O_2 + X \rightarrow R - \underset{\diagdown}{C} = O + H_2O$ $+ 37$ kg cal (3A)

may also take place to continue the chain, followed by either

$RCH_2OOH \rightarrow \underline{RCHO} + \underline{H_2O}$ (4) probably + ve

or

$RCH_2OOH \rightarrow RCH_2O - + OH$ $- 48$ kg cal (5)

with

$RCH_2O - + O_2 \rightarrow R - O - O + \underline{CH_2O}$ $+ 22$ kg cal (6)

When OH attacks the proximate RCH_2, either before or after disruption from RCH_2OOH, the olefine with the same number of C atoms is formed.

$OH + RCH_2 \rightarrow \underline{H_2O} + \underline{R'CH = CH_2}$ $+ 72$ kg cal (7) but the reaction

---

* *Cf.* Rice, 'J. Amer. Chem. Soc.,' vol. 56, pp. 488, 2105 (1934).

$$OH + RCH_3 \rightarrow \underline{H_2O} + RCH_2 \qquad + 18 \text{ kg cal} \quad (5\text{A})$$

takes place in the presence of sufficient fuel and thus leads to an increase in the number of radicals, and under suitable circumstances to ignition.

A feature of particular importance is that the process of chain-branching, which involves equation (5), leads to an absorption of about 48 kg cals, and will probably not take place unless the energy-rich peroxide can collide with a fuel molecule sufficiently quickly to include process (5A) in the energy balance. This characteristic feature of hydrocarbon combustion has already been commented on. Competing with this mechanism of chain-branching is the internal rearrangement of the peroxide RCH_2OOH to give RCHO, as in equation (4) (*cf.* Medwedew and Alexejewa for the evidence of this rearrangement*). If a rise in temperature favours this rearrangement at the expense of process (5A), which, in the limit for a very short-lived peroxide practically involves a triple collision $RCH_2OO + 2RCH_3 \rightarrow RCH_2O + 2RCH_2 - 20$ kg cal, the small negative temperature coefficient already referred to may be due to less chain-branching at higher temperatures (further evidence of the –ve temperature coefficient has recently been announced by Neumann and Aivazov).†

Finally, the RCH_2O — radical leads to the appearance of formaldehyde, as in (6),‡ which may also be formed from the oxidation of the methyl radical formed in (1A). A number of these steps requires activation energies, which cannot be computed at present, but the following information serves as further support.

i—*Initiation*

The difficulty of initiating hydrocarbon combustion has already been referred to. The above scheme involves initiation by aldehydes, which are probably formed heterogeneously, starting with a pure hydrocarbon oxygen mixture, and also involves a dehydrogenation of the hydrocarbon by means of peroxides, or the OH groups, etc., formed by rupturing them. Other initiating agents have been referred to above. The compounds which have been found effective in initiating the slow combustion of ethane (aldehydes, alcohol, water, iodine, and NO_2) probably lead to similar dehydrogenations.

The mechanism proposed very recently by Norrish§ for the oxidation of methane is in formal disagreement with that discussed here, since it involves the intervention of oxygen atoms and stresses the role of the methylene radical. Reasons have been given above for supposing that oxidation of the higher hydrocarbons over the range of temperatures from 200°–350° C differs from that at higher temperatures, at any rate in the relative importance of different steps in the combustion chain. The extension of a plausible mechanism for the oxidation of methane to cover the complex phenomena shown by the higher hydrocarbons does not seem to be justified.

---

* 'Ber. deuts. chem. Ges.,' vol. 65, p. 133 (1932).

† 'Nature,' vol. 135, p. 655 (1935).

‡ *Cf.* Rice, 'J. Amer. Chem.,' vol. 57, p. 350 (1935).

§ 'Proc. Roy. Soc.,' A, vol. 150, p. 36 (1935).

The following objections to an O chain may be mentioned—

(1) The abundant production of OH radicals in the flame suggest their important role in combustion.

(2) The production of alcohols by a mechanism

$$RCH_2O + RCH_3 \rightarrow RCH_2OH + RCH_2 \qquad + 10 \text{ kg cals.}^*$$

is necessarily favoured at high pressures when the peroxide is short lived.

(3) The initiation of chains by the production of O atoms from aldehyde peroxides seems less likely, owing to the energy requirements, than initiation by the OH produced by the rupture of an $-O-O-$ link. Furthermore, the same negative catalysts inhibit both aldehyde and (low-temperature) hydrocarbon oxidation, and the intervention of oxygen atoms in aldehyde oxidation seems very unlikely.

(4) The phenomena discussed with reference to the point of attack, etc., of hydrocarbons can hardly be explained if oxygen atoms are effective.

Without prejudicing the possible application of the low-temperature chain involving oxygenated radicals and compounds to high temperatures, it must be emphasized that theories giving a formally satisfactory explanation of the oxidation of methane, ethane, propane, and ethylene at high temperatures do not exhaust the complex phenomena observed for the higher hydrocarbons at low temperatures.

ii—*Processes of Initiation by Pyrolysis*

(*a*) $O_2 = 2O$—The equilibrium concentration of atoms at 523° K may be calculated from the data of Johnston and Walker.† When the ratio O_2/fuel is 1 : 1, there is one collision with an O atom for every 10^{21} collisions with an oxygen molecule. At ordinary pressures, the number of collisions is approximately 10^{10}/sec, so that even if every collision with an O atom led to a chain length of 10^5, the combustion would not be complete in less than 10^6 sec, which is far too long to account for the experimental results.

---

* Rice, 'J. Amer. Chem. Soc.,' vol. 57, p. 350 (1935).

† 'J. Amer. Chem. Soc.,' vol. 55, p. 187 (1933).

At 700° K the fraction of oxygen dissociated is 10^{-16} and the initiation in the gas phase by O atoms is perhaps feasible. The dissociation $H_2O = H + OH$ would lead to a sufficient concentration of OH for initiation at somewhat lower temperatures, and may explain the effect of water on the induction period.

(*b*) $C_3H_8 = C_3H_6 + H_2$—The thermal dissociation into olefine and hydrogen, of which propane is an example, has been invoked in some theories of combustion as the process of initiation.* At 523° K, using the data of Frey and Huppke†, a calculation shows that the equilibrium concentration of propene is about 0·2% at one atmosphere. The velocity with which the monomolecular dissociation takes place can be calculated from the data of Paul and Marek,‡ and even at 623° K the velocity constant is only 10^{-10} moles/sec. The large amounts of propene (up to 99% of the fuel) isolated in the slow combustion of propane by Pease§ can hardly be accounted for either from equilibrium or velocity considerations as due to pyrolysis, but is probably a stabilized by-product in some step in the chain, such as (7).

With the higher hydrocarbons, the amount of pyrolysis is greater, but branched isomers which show the greatest reluctance to combustion are the most dissociated. It does not seem as if dissociation into olefine and hydrogen can be the principal factor in initiating hydrocarbon combustion, especially as the olefines themselves show induction periods.

(*c*) $RCH_3 = RCH_2 + H$—Although data are wanting for a calculation of the equilibrium for this pyrolysis, the energy requirements are so large (94 kg cal) that similar arguments apply to those for the dissociation of oxygen. Furthermore, the dissociation $ROH = R + OH$ only requires about 72 kg cal, so that if pyrolysis is effective in initiation the alcohols should lead to it more readily than the paraffins. This does not seem to agree with the relative ease of combustion of paraffins and alcohols at 523° K, though it may, perhaps, hold for the high-temperature region.

iii—*Energy Requirements of the Various Steps*

Although it is not yet possible to estimate the activation energy required in the different processes in the scheme outlined, it is possible to check the actual energy changes, making use of known values of the bonding energies.||

---

* Lewis, 'J. Chem. Soc.,' p. 1555 (1927) ; p. 759 (1929).

† 'J. Ind. Eng. Chem.,' vol. 25, p. 54 (1933).

‡ 'J. Ind. Eng. Chem.,' vol. 26, p. 2034 (1934).

§ Pease, 'J. Amer. Chem. Soc.,' vol. 56, p. 2034 (1934).

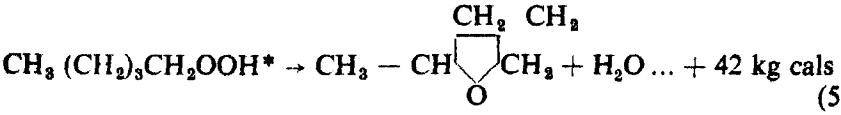
|| *Cf.* 'Ann. Rep., Chem. Soc.,' p. 401 (1931) ; Hinshelwood and Williamson, "Reaction between Hydrogen and Oxygen" (Oxford, 1934).

The figures given are based on the data: $H_2 = 2H - 101$ kg cal; $H_2O = H + OH - 112$ kg cal; $H_2O_2 = 2OH - 48$ kg cal; $RCH_3 = RCH_2 + H - 94$ kg cal; rupture of the $C - O$ link $- 72$ kg cal (average); $O_2 = -O-O- 59$ kg cal (half the dissociation energy into atoms, probably an over-estimate); $OH = H + O - 104$ kg cal; rupture of $C - C$, $- 71$ kg cal.

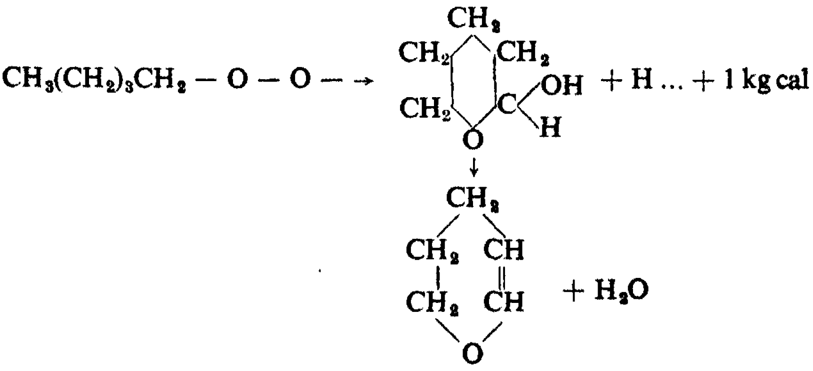
The rupture of organic peroxides across the $-O-O-$ bond is assumed to require the same energy as for hydrogen peroxide. From (ICT) data $C_3H_8 = C_3H_6 + H_2 - 21$ kg cals, so that $RCH_2 =$ Olefine $+ H - 28$ kg cal.

iv—*Further Evidence*

The numerous variants of the above scheme, which arise if side reactions are taken into account, complicate the study of the kinetics of oxidation of the higher hydrocarbons, which is not attempted here. The main purpose of the present discussion is to advance evidence for the principal chain-carriers in combustion. Apart from the facts already quoted, other evidence is available from the slow combustion of the higher hydrocarbons, since in this case the collision of an oxygen-rich carrier with a second fuel molecule may take place intramolecularly, leading to ring closure. To take one example, evidence is discussed in Part II which substantiates the formation of compounds such as tetrahydromethylfuran and dihydropyran in the combustion of *n*-pentane, these compounds being formed by intramolecular analogues of (5A) and (3) respectively, *i.e.*,



(5A′)



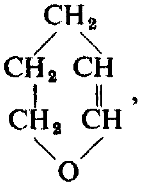
(3′)

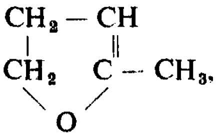
* 'J. Ind. Eng. Chem.,' vol. 26, p. 2034 (1934).

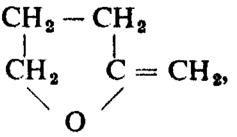
The energies are calculated assuming the average values for the links C — H, C — O, and O — H already used, and thus neglect any strain in the rings. Direct combustion of tetrahydropyran which might be formed as in (5A′) does not seem to give rise to dihydropyran, whose formation thus requires a different mechanism (3′).

Such intramolecular oxidations are favoured by the fact that the oxygen-rich carrier will usually have a high vibrational energy-content as the result of its formation, without requiring to meet a second fuel molecule in the same condition. A similar process probably explains the results obtained in the slow oxidation of paraffins with very long chain, *e.g.*, $C_{30}H_{62}$, as studied by Francis and others.* For such paraffins, oxidation of the fourth carbon atom from the end is a characteristic feature, and is probably due to attack by the oxygen-rich end group, with temporary or permanent ring enclosure.

The great reactivity of dihydropyran and the epoxypentenes (which may likewise be found)







with oxygen, water, acids, aldehydes, etc., gives some explanation of the very complex mixture of products obtained in the slow combustion of the higher hydrocarbons (*see* Part II).

SUMMARY

In a discussion of some of the phenomena of hydrocarbon combustion, reference is made to the sources of information on the reaction mechanism. These include the chemical identification of intermediates of oxidation, reaction kinetics, chemi-luminescence, the induction period and wall effects, the action of positive and negative catalysts, and a comparison of the behaviour of isomers, and of substituted hydrocarbons. The probable importance of vibrational energy, particularly at low temperatures, is substantiated, using simple models of the molecules. A mechanism of reaction based on this discussion is put forward. This involves peroxide radicals and molecules, and in particular provides a mechanism of chain-

* Francis and Wood, 'J. Chem. Soc.,' p. 501 (1922) ; p. 1897 (1927).

branching which accounts for special phenomena observed in the low-temperature region of combustion. Further support for the mechanism is obtained from the oxidation of hydrocarbon molecules sufficiently long to permit ring closure.

Investigations on the Combustion of Hydrocarbons II—Absorption Spectra and Chemical Properties of Intermediates

By A. R. UBBELOHDE, M.A. (Oxon), Department of Thermodynamics, Clarendon Laboratory, Oxford

(*Communicated by A. C. G. Egerton, F.R.S.—Received May* 24, 1935)

[PLATES 5 and 6]

1—SPECTROGRAPHIC INVESTIGATIONS OF POSSIBLE INTERMEDIATES

The investigations recorded here are a development of the work of Egerton and Pidgeon* on the absorption spectra of burning hydrocarbons, which had included investigations of the absorption and slow combustion of alcohols up to amyl alcohols, aldehydes up to valeraldehyde, acids up to butyric; amylene, ethylene, ethyl acetate, ethyl hydroperoxide, diethyl peroxide, and of anti-knocks such as lead tetraethyl, iron carbonyl, butyl iodide, and aniline.

The results were compared with the slow combustion of the normal hydrocarbons up to pentane, in the same apparatus. The only intermediates which could be demonstrated spectroscopically in the slow combustion of hydrocarbons were formaldehyde and organic acids, mainly because only these substances have a sufficiently large absorption coefficient to be detectable at the concentrations occurring. In addition, a characteristic band was discovered at the end of the (pseudo) induction period, in the slow combustion of the higher hydrocarbons, though the molecule responsible for it was not identified, in spite of a considerable search among the ordinary products of slow combustion.

Since this band was apparently common to a number of hydrocarbon oxidations, it was obviously important to investigate its nature in a special research. In the experiments to be described in this paper, the following investigations were pursued:

---

* 'Proc. Roy. Soc.,' A, vol. 142, p. 26 (1933).

(1) In order to map the possibilities of the method, a study was made of all the compounds with characteristic absorption spectra that might arise in slow combustion. The compounds investigated are listed on the chart. Ethyl ether was also reinvestigated on account of its very easy peroxidation.

(2) The influence of certain catalysts such as NO_2 on the slow combustion of hydrocarbons was studied, to see whether any fundamental difference of process was detectable.

(3) A special study was made of the absorption spectra and slow combustion of ring compounds, including one O atom, in an attempt to discover the origin of the band arising in the slow combustion of the higher hydrocarbons. Since this first appears for butane, it might seem to depend on ring formation of some kind.

Among the compounds investigated were furan, furyl alcohol, furfuraldehyde, furoic acid, succinic anhydride, α-methyl furan, tetrahydromethyl furan, epoxy 1-4 pentene 4, and pentene 5, tetrahydrofuryl alcohol, butyrolactone, α- and β-angelilactones, γ-valerolactone, levulinic acid, ethers of tetrahydrofuryl alcohol, dihydropyran, tetrahydropyran, and dimethyl pyrone.

(4) Spectroscopic methods were repeatedly used in investigating the nature of the products of slow combustion obtained as described below, and in particular in studying the properties of the substance X, responsible for the new absorption band, once its method of isolation had been worked out.

EXPERIMENTAL METHODS AND RESULTS

The experimental methods have already been described.* Air, or when necessary nitrogen, was saturated with the vapour of the substance at various temperatures, and the absorption spectrum of a column about 80 cm long was measured with an ultra-violet spectrograph Hilger E.31; the spectra extended from 4200 A to 2000 A with the high-frequency discharge in hydrogen as source. When necessary the column could be shut off, and its temperature gradually raised to study the slow combustion of the contents. In some cases the product was carefully pushed along the tube with nitrogen, in order to study the absorption of different lengths of column. A summary of the regions and general types of absorption

---

* Egerton and Pidgeon, 'Proc. Roy. Soc.,' A, vol. 142, p. 26 (1933).

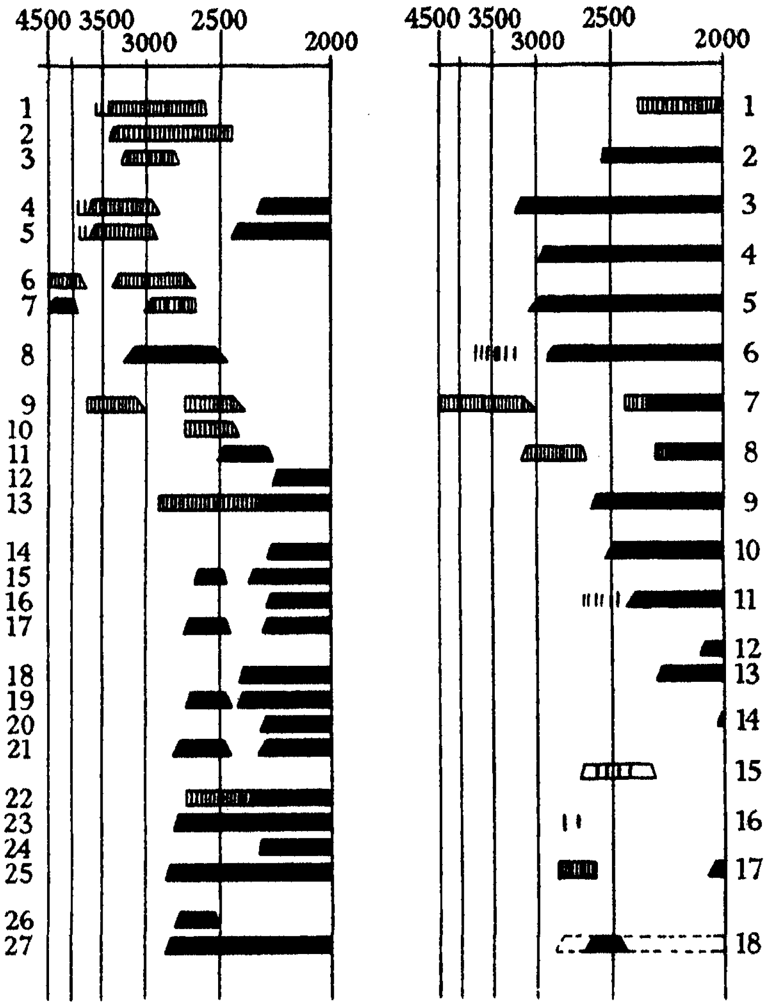


Fig. 1.

obtained under headings (1) and (3) is made in the form of a chart. This represents the general regions of absorption, with the following conventions:

(1) The concentration of vapour was selected so as to show clear absorption spectra which could be re-identified, if necessary, in a mixture of unknown substances, since the temperature of such a mixture could be varied till it likewise gave a suitable concentration of vapour. The chart and spectra were thus used for the qualitative identification of substances under circumstances which made chemical identification impossible.

FIG. 1—*Description of chart of general regions of absorption in the ultra-violet.* Band absorption is indicated by fine lines, continuous absorption by black bands. The selection of suitable concentrations has been referred to in the text. Wave-lengths were based on photographs of the mercury arc.

(The gaps correspond to gaps in the chart)

(1) Formaldehyde
(2) Acetaldehyde
(3) Butyraldehyde

(4) Acrolein
(5) Crotonaldehyde

(6) Glyoxal
(7) Diacetyl

(8) Ketones

(9) Furfuraldehyde
(10) Furyl alcohol
(11) Furoic acid
(12) Furan
(13) Methyl furan

(14) Tetrahydromethylfuran
(15) *Id.* (burning)
(16) Tetrahydrofuryl methyl ether
(17) *Id.* (burning)

(18) α Angeli-lactone
(19) *Id.* (burning)
(20) β Angeli-lactone
(21) *Id.* (burning)

(22) Dihydropyran
(23) *Id.* (burning)
(24) Tetrahydropyran
(25) *Id.* (burning)

(26) X when dilute
(27) X when strong

(1) Formic acid

(2) Dimethyl keten

(3) Acetyl acetone

(4) Ethyl acetoacetate

(5) Diethyl and ethyl hydroperoxide

(6) Amyl and ethyl nitrites

(7) Nitrogen peroxide

(8) Sulphur dioxide
(9) Dihydromethylfuran

(10) Epoxy 1 · 4 pentene 5

(11) Penten 5–1 ol

(12) Amylene
(13) *Id.* (burning)

(14) Ethylene

(15) Benzene

(16) Phenol

(17) Naphthalene

(18) Methyl iodide

The vertical lines correspond to the wave-lengths 4500, 4000, 3500, 3000, 2500, and 2000 A, from left to right.

(2) Regions of banded absorption are indicated by fine vertical lines, but the exact positions are only given for substances like benzene, where the chief bands are very prominent. Continuous absorption is indicated by black bands, the slope of the edge being a measure of the sharpness of the end of the absorption region.

The following notes may be added with regard to the behaviour of the compounds in the cold, and on slow combustion.

(*a*) All compounds containing a —CHO group showed a series of absorption bands roughly in the same region (3000–3600 A), but minor differences made it quite easy to distinguish the aldehydes. Thus formaldehyde and aldehydes with a double bond conjugated with C=O have sharp band spectra, quite distinct from that of acetaldehyde, and still more from that of the higher aliphatic aldehydes, which have diffuse bands only.* In addition, glyoxal and diacetyl showed a second region of absorption between 4500 A and 4000 A. Acetal in agreement with its chemical inertness showed no marked absorption. Furfuraldehyde in addition to the strong absorption of the furyl group showed weak absorption in the aldehyde region. On slow combustion the aldehydes burnt away, and generally gave rise to formaldehyde.

(*b*) Unsaturated compounds gave a remarkably sharp cut-off, the position being nearer the visible in the higher olefines. This behaviour was common to all unsaturated compounds investigated, including allyl derivatives and furane. The energy involved in the absorption process corresponds roughly with that required to break one link of the C=C double bond; if this proved to be the process of electronic excitation, a valuable optical means of testing the influence of substituents on the double bond would be available. On slow combustion the olefines gave increased cut-off and formaldehyde, but no evidence was ever obtained of the formation of X.

(*c*) All the ketones CH_3COCH_3, $CH_3COC_2H_5$, $C_2H_5COC_2H_5$, and $CH_3COC_3H_7$ produced the same region of continuous absorption in the spectra (*cf.* fig. 15, Plate 6). This continuous band is closely similar to that of X in appearance, but the position of its maximum is definitely nearer the visible. On slow combustion up to 370° C tests with Schiff's reagent showed the production of aldehydes, but the amounts were too small for optical identification.

(*d*) Organic peroxides of all kinds gave continuous absorption which extended further to the visible at higher concentrations. It was interesting to note that with $C_2H_5OOCH_2(OH)$ formaldehyde appeared at 45° C, probably owing to thermal dissociation of the compound. With diethyl peroxide and air flowing through the heated tube chemical tests showed the production of CO_2 at 110° C, and optical tests showed HCHO at 170° C. With C_2H_5OOH corresponding temperatures were 130° and 200° C. These very low temperatures have already received comment.

---

* *Cf., e.g.*, Leighton and Blacet, 'J. Amer. Chem. Soc.,' vol. 55, p. 1766 (1933).

(*e*) Benzene showed its characteristic absorption spectrum, which was modified by substituents as in phenol and naphthalene. The absorption coefficient is so high that the four lines 2650–2420 A were sometimes found in quite distinct products boiling around 80° C. Benzene oxidized to give a sharp cut-off.

(*f*) The absorption spectrum of SO_2 was included since it was formed, *e.g.*, in the oxidation of impure ether. Methyl iodide had an absorption region almost identical with X. NO_2 is formed in many combustions, but only at temperatures and pressures higher than those used in these experiments. Special experiments, fig. 5, Plate 5, showed that although the addition of NO_2 lowered the temperature of first appearance of X with hexane from 300° to 240° C, there was no other difference detectable in the slow combustion. Amyl and ethyl nitrites had similar absorption spectra, fig. 16, Plate 6, and the addition of 0·03% of ethyl nitrite to hexane lowered the temperature of appearance of X to 255° C.

(*g*) A specially purified sample of *iso*-pentane was tested to see if branched isomers gave X, and a very definite band was noted. Dimethyl keten gave a sharp cut-off, but no absorption corresponding with a carbonyl group, in agreement with the chemical properties. The combustions of propyl and *iso*-propyl alcohols gave no X bands or features of interest apart from the formation of formaldehyde. Acetyl acetone and ethylacetoacetate gave sharp cut-offs, and acetic anhydride a small cut-off only.

(*h*) Furan gave a sharp cut-off, which moved towards longer wavelengths as the furan air mixture (of constant composition) was raised to 100° C. No combustion products were optically detectable up to 250° C, fig. 12, Plate 6. On introducing a methyl group, as in α-methylfurane, a set of sharp bands appeared at higher concentrations, fig. 13, Plate 6, in addition to a similar cut-off. The substitution of either $—CH_2(OH)$ or —CHO in the furan nucleus gave a very characteristic set of bands, fig. 6, Plate 5, the absorption coefficient being so high that furyl derivatives are readily shown, *e.g.*, in air passed over gently warmed sucrose. Substitution of —COOH gave a region of continuous absorption in place of the sharp furyl bands, and a shift to the ultra-violet, indicating an interaction between the —COOH and the furyl nucleus, in agreement with the chemical properties. Succinic anhydride gave practically no absorption, suggesting the absence of the enol form.

On slow combustion all these compounds just burned away, the absence of characteristic intermediates being partly due to the fact that once the ring is broken very easily oxidized substances are formed, and do not accumulate, and partly to the high-absorption coefficient of furyl compounds, which necessitated the use of low concentrations, thus reducing the chance of observing oxidation products of low-absorption coefficient.

(*i*) Tetrahydromethylfuran and tetrahydropyran gave a sharp cut-off which was likewise present in a number of derivatives. On slow combustion, a characteristic cut-off and the production of formaldehyde bands was observed, but the absorption in the far ultra-violet was too complete to detect if X was formed, figs. 10 and 11, Plate 6.

Epoxy 1·4 pentene 4 and pentene 5 gave larger cut-offs, and similar absorption spectra on slow combustion as the fully hydrogenated compounds. Dihydropyran gave a very characteristic band spectrum in addition to the cut-off of O ring compounds, fig. 14, Plate 6; on combustion all these compounds resemble the hydrocarbon spectra, except that absorption in the far ultra-violet is too complete to show X bands, if they were formed. The alcohol 1 hydroxy pentene 5 gave a cut-off, but no marked phenomena on combustion.

Tetrahydrofuryl alcohol is so involatile that no marked absorption was observed with the vapour up to 80° C, but the methyl and ethyl ethers gave the same cut-off as tetrahydromethylfuran, and on slow combustion gave rise to an absorption band closely resembling X, but probably slightly further to the ultra-violet, fig. 7, Plate 5.

Pure ethylene oxide and dimethyl pyrone vapour up to 80° C showed no marked absorption, nor did levulinic acid, γ-valerolactone, and butyrolactone. Both α- and β-angelilactones showed cut-offs, and on combustion gave a continuous absorption band closely resembling X, fig. 8, Plate 5.

The conclusions from these investigations were that the characteristic X band was produced by normal hydrocarbons from butane upwards, and by *iso*-pentane. Ethyl nitrite and NO_2 lower the temperature of its first appearance, but not the general course of its formation. No X band was produced by butylene and the amylenes, nor did the interruption of the butane chain by oxygen, as in ethyl ether, lead to the production of X, in spite of the much lower temperature of combustion.* A cut-off and formaldehyde bands produced by tetrahydromethylfuran, tetrahydropyran, and a number of derivatives on slow combustion resembled the general phenomena for hydrocarbons, but if X was produced its concentration was too high for the short wave-length edge of the absorption region to be visible. Regions similar to X were, however, produced in the slow combustion of ethers of tetrahydrofuryl alcohol, and α- and β-angelilactones. None of the stable compounds investigated (except methyl iodide) showed absorption corresponding to X, so that the molecules responsible for it are not included in the stabilized products of combustion hitherto reported.

---

* Absorption only occurs in the formaldehyde region, fig. 17, Plate 6.

X is thus a combustion product of the paraffins, and possibly of a number of O ring compounds, such as the unsaturated lactones and tetrahydrofuryl ethers.

2—CHEMICAL INVESTIGATION OF INTERMEDIATES

The results described in the previous section suggested a chemical study of intermediates in the slow combustion of hydrocarbons. In pentane, intermediates formed at the *beginning* of the attack on the molecule would mostly still contain five carbon atoms, with the possible exception of formaldehyde. Following the spectrographic indications of the previous section, the chemical evidence for the formation of O ring compounds was investigated. Special attention was also paid to the peroxides which could be isolated during slow combustion.

The outcome of these investigations has been to substantiate both the formation of fully hydrogenated rings and of dihydropyran, etc. The peroxides which could be isolated were at least in part formed by the direct addition of oxygen to unsaturated O ring compounds. Furthermore, the substance X responsible for the absorption band in the slow combustion of paraffins arises from the decomposition of these peroxides, and apparently resembles the hydrofuryl residue of ascorbic acid, both in the absorption spectrum, large absorption coefficient, and chemical instability.

Preliminary experiments were performed on the slow combustion of specially pure hexane, using a method similar to that described by Mondain Monval and Quanquin.* At 300° C the combustion products separated into a hydrocarbon layer and a dense aqueous layer, both smelling of burnt sugar, aldehydes, etc. The aqueous layer had a much stronger peroxide reaction than the hydrocarbon layer, and the vapours from the aqueous layer likewise showed a much stronger X band, thus showing for the first time that this substance was sufficiently stable to be isolated, at least in a mixture. Fractionation of the hexane layer gave a fraction boiling 75°–85° with a sharp ethereal smell characteristic of O ring compounds. One experiment gave furfuraldehyde derivatives with very definite aniline acetate reaction and furyl absorption spectrum, but this could not be repeated.

---

* 'Ann. Chim. (Phys.),' vol. 15, p. 309 (1931).

Subsequent experiments were carried out with *n*-pentane, for the following reasons.

(*a*) Pentane boiling at 36·5° could be freed with certainty from all naphthenes and aromatic compounds, as well as from olefines. This was an important consideration, since the large amounts required in view of the complex mixture of products precluded the use of synthetic hydrocarbons. The chief impurity in the pentane used was *iso*-pentane, and special experiments with pure *iso*-pentane showed that although its slow combustion was less marked that that of *n*-pentane under the same conditions, the same band X was produced at the end of a (pseudo) induction period. Furthermore, combustion of the *iso*-compound takes place more readily in the mixture than when pure.

(*b*) The rather complicated chemistry of some O ring compounds with five carbon atoms has recently been worked out by Paul,* so that comparison compounds were more readily available than with hexane.

(*c*) Although the stability of certain important products probably increases with the number of carbon atoms, the number of possible products likewise increases to an undesirable extent.

Tests showed that a moderate excess of fuel favoured the production of peroxides, though a large excess seems to suppress them. In order to economize pentane, a circulatory system was constructed, so as to use the pentane again after removing high-boiling products of combustion. The working of the system is clear from fig. 2. A stream of air from an electric blower (H_2SO_4 drying made no difference) is passed through pentane in the carburettor A, the temperature being maintained electrically. The fuel-air mixture enters the Pyrex combustion chamber at B; the temperature is controlled electrically (320°–350°) and is measured externally with a thermocouple, the adjustment of temperature and rate of flow, so as to give a good production of peroxide, being empirical. A Pyrex glass inner tube was added in some experiments and was cooled with water, but the additional cold surface did not appreciably increase the yield of peroxide.

Observation of the tube in the dark showed a moderate chemiluminescence near the exit end when the peroxide was being produced at a suitable rate. The products of combustion were passed into an expansion chamber C, which was filled nearly completely with Pyrex glass condenser tubing, through which cold water circulated. The condensate was collected in D, and the lower layer was drawn off from time to time. The upper layer and a further condensate collected in a liquid air trap were periodically returned to the flask F, where fractions boiling below 37°, and containing pentane, unsaturated substances, formaldehyde, and acetaldehyde, were returned to the carburettor by distillation. The higher boiling fractions gradually accumulated in F. Fresh pentane could be added at G. With this apparatus 10–22 cc of crude aqueous peroxide were drawn off per hour, and 150 cc of pentane used up.

---

* 'Bull Soc. chim. Fr.,' vol. 53, p. 417; vol. 54, p. 1489 (1933); vol. 1, pp. 971, 1397 (1934).

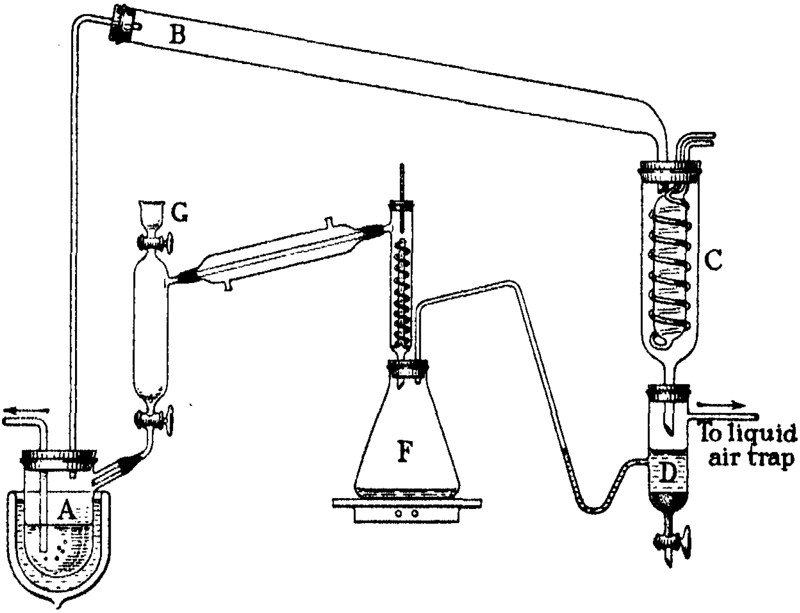


FIG. 2.

As a result of repeated use of the apparatus the products of combustion gradually accumulated in the higher boiling fractions in F, and in the aqueous layer which was drawn off from time to time. Owing to the use of a circulatory method, estimates of the relative amounts of the various products were not obtained, the method being designed for the collection of the higher boiling fractions only. The usual preliminary tests on organic compounds showed the presence of very reactive products, which gave strongly coloured compounds even with reagents such as concentrated HCl; only a simplified account of the treatment of the mixture is given here, since other methods used did not lead to the separation of characteristic compounds.

Owing to the fact that the water-soluble substances could not be readily distilled from the aqueous layer without destroying the bulk of the peroxide, isolation of O ring compounds was undertaken chiefly with the residue accumulating in F. On careful fractionation, this gave, amongst other products, a fraction boiling at 70°–90° C, which contained cyclic O compounds. This was established as described below.

Tests for Cyclic O Compounds in the Pentane Layer—Fractionation of the residues in F gave large amounts of hydrocarbon, etc., boiling up to 50° C. The absorption spectrum showed the presence of a substance giving rise to bands similar to dihydropyran even in these low-boiling fractions, and treatment with bromine followed by distillation showed that about one-third of the mixture was unsaturated. The bromides decomposed on boiling at atmospheric pressure, suggesting the presence of furan and possibly pyran derivatives.

The temperature then rises rapidly to 63°, and after a small arrest at 73° the bulk of the fraction 65°–95° boils in the neighbourhood of 80°. The absorption spectrum shows dihydropyran, particularly in the fraction 81°–105°. The residues boiling above 105° contained no readily identified products of combustion, apart from aldehyde polymers.

The fraction boiling 65°–95° had a sharp ethereal smell, together with some more acrid impurity. On heating with acetic anhydride, only very small amounts of acetates were formed. (The bulk of such alcohols as are formed would remain in the aqueous layer.) On adding bromine vigorous absorption took place. When the yellow colour persisted for some time, the product was repeatedly washed with water to remove cyclic O compounds, etc., and the aqueous layers were salted out with K_2CO_3. Water distilled from the K_2CO_3 waste was used again for economy. By this means about half the original 65°/95° fraction was recovered as a liquid soluble only in a fair excess of water, lighter than water, and almost free from unsaturated compounds. The greater part boiled 75°/85° and had a smell of O ring compounds. The absorption spectrum both in the cold (cut off) and on slow combustion (formaldehyde bands, etc.) was similar to methyltetrahydrofuran and tetrahydropyran, fig. 9, Plate 6. The refractive index of 1·392 at 16° compares with 1·406 for synthetic methyltetrahydrofuran at 14°, showing that the distillate 75°/85° was not quite pure. The chief impurity was probably ethyl alcohol, since on distilling with acid $K_2Cr_2O_7$ a little acetaldehyde was obtained, detected from its absorption spectrum.

An attempt at further purification was made by boiling the crude product as salted out with K_2CO_3 with four volumes of 60% HBr on a water-bath, using a reflux condenser. On first adding HBr, the mixture blackened and became very warm; this behaviour was strongly reminiscent of partly hydrogenated furan or pyran derivatives, suggesting incomplete removal. A small amount of methyl bromide was evolved. Steam distillation of the mixture after 2 hours' heating gave about 6% of the original ring compound unchanged, together with bromides, 75% of which distilled through the range 180°–210° (Lipp* gives 198°/201° for $CH_3CHBr . CH_2 . CH_2 . CH_2Br$ and $(CH_2)_5 Br_2$ has the boiling-point 220°). A second distillation gave about 35% boiling 198°–201°, corresponding with methyltetramethylene bromide. The substance was not pure, however, since it rapidly coloured blue, then brown in air, and resinified a little. These colours suggested the presence of bromine derivatives of partly hydrogenated furans and pyrans, which might have escaped bromination, or have dissolved in the water used for washing out the cyclic O compounds.

Analysis showed that the percentage Br was only 60% in lieu of 69·5%, the approximate composition of the impure brown liquid being $(CH_2)_5Br_{1\cdot5}O_{0\cdot5}$. In order to remove oxygen compounds, the mixture was dissolved in pure pentane, and shaken with strong sulphuric acid till the acid layer no longer gave a cloud on pouring into water. After drying the pentane layer with $CaCl_2$, the compound on evaporating the pentane no longer coloured rapidly in air, and the percentage Br, as determined by direct interaction with alcoholic $AgNO_3$ (the action is fairly rapid), had risen to 66%. The portion soluble in H_2SO_4 was extracted with pentane from the water, and after drying still reacted with alcoholic $AgNO_3$, but much more slowly. It may have contained monobrompyran from the distillation of dibrompyran.†

The bromides left after washing out cyclic O compounds from the brominated 65°/95° fraction with water, gave on distillation a number of fractions, chiefly boiling below the major fraction 95°–110°, with fuming and darkening in air; an important constituent was probably monobrompyran from the decomposition of dibrompyran, but the presence of bromofuran derivatives is not excluded, and made the mixture inconveniently complicated for working up the small amounts available.

If bromination of the crude 65°–95° fraction was omitted before treatment with saturated HBr, the dibromides distilled as before, but the pyran derivatives were almost completely lost owing to their solubility in water, resinification in the presence of acids, etc.

---

* 'Ber. deuts. chem. Ges.,' vol. 22, p. 2569 (1889).

† Paul, 'Bull. Soc. chim. Fr.,' vol. 53, p. 417; vol. 54, p. 1489 (1933); vol. 1, pp. 921, 1397 (1934).

To sum up the investigation of the pentane soluble residues, the presence of considerable amounts of saturated ring compounds such as methyltetrahydrofuran, and probably tetramethylene and pentamethylene oxide, was inferred from the boiling-point, solubility in water, absorption spectrum in the cold and on burning, refractive index, and formation of bromides of the right boiling-range, though these have not been obtained absolutely pure, on account of the rather small amounts of the mixtures available. The presence of dihydropyran was inferred from the very characteristic absorption spectrum, violent bromination, and behaviour of the dibromide on distillation. The chemical properties of the mixtures containing dihydropyran are likewise in agreement with the published properties of this very reactive substance.

Hydroxyacids, aldehydes, alcohols, etc., were also detected in other fractions either from the residues in F, or dissolved in the aqueous layer, but were not obtained sufficiently pure from the complex mixture for complete identification. The amount of aldehydes formed was very considerable, but only traces of ketones could be detected. This has been commented on in Part I, and suggests that the oxidation of end groups preponderates.

With hexane analogous compounds were formed even more readily, but have not been sufficiently investigated in the literature to make their detection easy. Absorption spectra of compounds similar to dihydropyran were observed.

Tests on the Peroxide

Three slightly different specimens of a peroxide isolated during the slow combustion of pentane were available: (*a*) A small concentration in the pentane layer, which was thus preserved from hydrolysis by acids, etc. (*b*) A solution about 1·5 molar with respect to peroxide was drawn off from time to time during the slow combustion of pentane. This aqueous layer was strongly acid, and contained aldehydes, O ring compounds, dihydropyran etc., which would tend to modify the original peroxide by hydrolysis, formation of hydroxyalkylperoxides with the aldehydes present, etc. (*c*) A small amount of peroxide was collected as a pentane insoluble aqueous layer in the liquid air trap, so that the changes described under (*b*) would be arrested.

Owing to the larger amounts available, the bulk of the experiments were carried out with (*b*). On heating with dilute acids this solution went turbid and charred. On distillation either at ordinary pressures or at a pressure of 6 mm of mercury a very complex series of fractions was obtained. At first there was marked frothing and the evolution of considerable amounts of acetaldehyde, accompanied by the destruction of the bulk of the peroxide. Alkylene oxides, dihydropyran, etc., then distilled and finally at ordinary pressures a substance distilled with water from 95°–98° C, the vapour from which gave a very much stronger X band than usual. This fraction had no peroxide properties ($B = F = M =$ nil, *see* below) and is referred to again.

At a pressure of 6 mm of mercury, similar results were obtained, but in addition small amounts of a peroxide (*d*) were collected from 60°–100°, the bulk distilling 80°–100°.

Chemical Tests

A number of reagents useful for the qualitative differentiation of organic peroxides has been described.* The most important of these are acid KI+ starch alone (B) compared with the same reagent with a trace of ferrous salt (F), or a drop of ammonium molybdate under CO_2 (M). A further test useful in the presence of excess of unsaturated compounds which absorb iodine is ammoniacal lead hydroxide, which turns brown with a number of peroxides; the test can be intensified with tetrabase and glacial acetic acid, or KBr + fluorescein (lc).

Organic peroxides may be roughly classified by their behaviour with these reagents. With fairly dilute solutions (less than 10^{-2} molar) the results given in Table I are obtained.

These tests may be compared with the samples *a*, *b*, *c*, and *d* isolated from the slow combustion of pentane by the methods described. All these samples decolorize alkaline methylene blue, and in addition the peroxide in the crude aqueous layer evolves hydrogen with 50% KOH. Owing to the simultaneous presence of formaldehyde in this layer, and possible hydrolysis of peroxides to H_2O_2 by the acids present, this test does not necessarily indicate the *primary* formation of a peroxide $R.OOCH_2(OH)$. The peroxides *a* and *d*, though they decolorized alkaline methylene blue in the cold, gave no signs of evolving H_2 with stronger alkali.

With acid potassium iodide + starch, all these peroxides gave $M > B$, initially (F rapid but fades), and the unhydrolysed peroxide in the pentane layer (*a*) gave $M = B$ in 15 minutes, this behaviour suggesting an olefine peroxide.

Tests were also made of the effect of these peroxides on engine knock, since the work referred to above had suggested that only certain types of peroxides gave rise to knock. The pro-knock effect of the peroxide in the crude aqueous layer was abnormally low for an ROOR peroxide, and although it was concentrated by adding ether, removing water with anhydrous Na_2SO_4 and $CuSO_4$, and evaporating the ether again, the product still had only a moderate pro-knock effect. A further peculiarity was that the knock only set in after the substance had been sprayed into the throttle for about 300 revolutions. Valeraldehyde alcoxy peroxide of about the same oxidizing power gave rise to immediate definite knock.

---

* Egerton, Smith, and Ubbelohde, 'Phil. Trans.' A, vol. 234, p. 433 (1935).

TABLE I

Peroxide	Immediate	5 min	15 min	Remarks
H_2O_2	$M > F > B$	$M > F > B$	$M = B > F$	Strong H_2O_2 + M bleaches HI
C_2H_5OOH	$F > M > B$	$M > F > B$	$M > B > F$	—
C_3H_7OOH	$M > B$	$M > B$	$M > B$	Both *iso* and normal
C_4H_9OOH	$M > B$	$M > B$	$M > B$	Both *iso* and normal
$C_2H_5OOC_2H_5$	$F > M > B$	$F > M > B$	$F > M$	B practically nil even in 15 min
$(CH_3CO)OOH$	$M = B > F$	$M = B > F$	$M = B > F$	Action too rapid for catalysis by M
$(C_4H_9CO)OOH$	$M = B > F$	$M = B > F$	$M = B > F$	—
Benzoyl peroxide	$M > B > F$	(Action very slow owing to low solubility)		
Amylene peroxide	$F \gg M > B$	$M \gg B > F$	$M = B > F$	—
Dimethylketen peroxide	$M > B$	$M \gg B$	$M = B$	(F nil throughout)
$C_2H_5OOCH_2(OH)$	$M > B > F$	$M > B > F$	$M > B > F$	Decolorizes alkaline methylene blue (10% NaOH)
$CH_2(OH)OOH$	$M > B$	(F rapid, then fades)		Decolorizes alkaline methylene blue
$CH_3CH(OH)OOH$	$M > B$	(F rapid, then fades)		No action on alkaline methylene blue
$C_4H_9CH(OH)OOH$	$M > B$	(F rapid, then fades)		No action on alkaline methylene blue

Other peroxides tested for their influence on knock were ethylhydroperoxide, diethyl peroxide, butylene peroxide, dimethylketen peroxide, acetyl peroxide, butyryl peroxide, *iso*-propyl, propyl, and butyl hydroperoxides, and the alcoxy peroxides derived from H_2O_2 combined with HCHO, CH_3CHO, C_4H_9CHO, $C_5H_{11}CHO$. When sufficient quantities were available and the solubility in petrol was sufficient, standard solutions were made up, but in a number of cases the rougher test of spraying a fine vapour of peroxide into the throttle at a constant rate had to be used. Ethyl and amyl nitrites, ethyl nitrate, nitromethane, nitrobenzene, and nitrogen peroxide have also been tested,* and all fit in with the scheme of rupture already discussed, when pro-knock activity is evident at concentrations of about 10^{-5} mol fraction.

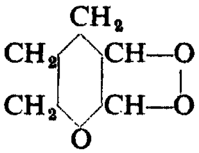
---

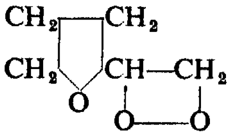
* *Cf.* Egerton, Smith, and Ubbelohde (*loc. cit.*).

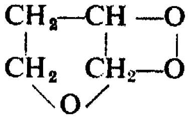
Thus the chemical tests (with KI) suggested the presence of an olefine type of peroxide, and also an alcoxy peroxide (with alkaline methylene blue). The concentrated peroxide had no corrosive action on the skin, and the reaction with $NaHSO_3$ or zinc + glacial acetic acid was not violent. In addition, the knock tests suggested an olefine type of peroxide.

A New Type of Peroxide: Vinyl Ether Peroxides

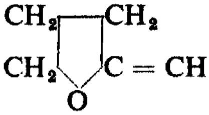
The properties of the peroxide isolated during the slow combustion of pentane agree in a number of ways with vinyl ether peroxides prepared from the three hydrogenated furans and pyrans

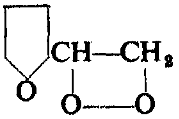




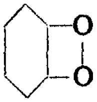
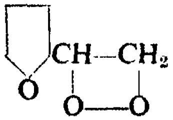


and since some of these compounds are formed during slow combustion, their further peroxidation at 300° C is almost inevitable. One source of the peroxides isolated during the slow combustion of the higher hydrocarbons is thus made clear.

All these vinyl ether peroxides were prepared by bubbling oxygen through the corresponding liquid, preferably in sunlight. Epoxy 1·4 pentene 5  peroxidizes so readily that on passing oxygen through the liquid near the boiling-point (80° C) a thick mist of peroxide was formed in the vapour, the effect being reminiscent of the mists formed during the slow combustion of pentane. The peroxide of dihydropyran is formed more slowly. When oxygen was slowly bubbled through the liquid at 60° C for 12 hours, using a reflux condenser, sufficient peroxide accumulated to give a very definite test with the lead reagent, and with alkaline methylene blue, but the yield was very poor. Attempts at preparing this peroxide at 200° C in pure oxygen led to an explosion and were discontinued. The rate of formation of epoxy 1·4 pentene 4 was intermediate between that of the two others.

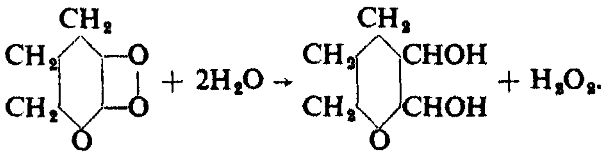
These vinyl ether peroxides combine the properties of alcoxy peroxides (to which they would give rise on opening the ring by hydrolysis) with those of olefine peroxides. Thus they all bleached alkaline methylene blue in the cold, the reaction being most rapid with the  peroxide.

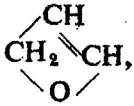
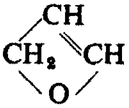
With acid potassium iodide the following behaviour was observed:

Peroxide	Immediate	5 minutes	15 minutes
	F > M > B	M > B (F fades)	M ≫ B ≫ F
	F > M > B	F = M > B	F = M ≫ B

The resemblance was thus closest between the peroxide of dihydropyran and that isolated as (*a*) or (*d*) in the slow combustion of pentane. The only difference from (*b*) observed was that dihydropyran peroxide did not give hydrogen with concentrated NaOH alone. Nevertheless, after standing with HCHO and acetic acid for an hour, dihydropyran peroxide evolved a gas in the cold on adding alkali, which was probably hydrogen. As has been explained above, only the peroxide (*b*) in the aqueous layer was observed to give hydrogen with strong alkali, and this may have been due to the liberation of H_2O_2 in the presence of formaldehyde, by hydrolysis.

The probable course of this hydrolysis* for an analogous reaction with olefine peroxides, leads to hydroxyaldehydes related to the pentoses; this may account for the smell of burnt sugar frequently observed during slow combustion, and near top dead centre in samples taken from the engine cylinder.



At present, the question whether other types of peroxide are formed as primary products of combustion, and not by subsequent reactions in the aqueous layer, must be left open. Two further points may be mentioned, however. An " alcoxy peroxide " was isolated during the slow combustion of propane by Pease.* An analogous formation of  propylene oxide would account for the unexplained deficit of carbon compounds mentioned by the author, but the chemistry of the compound  and its peroxide is too vague to know whether it could account for all the results observed. The similarity between the behaviour of the alcoxy and vinyl ether types of peroxide has already been pointed out.

---

* *Cf.* Yule and Wilson, 'J. Ind. Eng. Chem.,' vol. 23, p. 1254 (1931).

A second point of interest was the comparison of the total number of molecules in the crude and partially purified peroxide from pentane, with the total number of peroxide molecules, and with the average molecular weight. Experiments were carried out by measuring the freezing-point depressions of solutions containing weighed amounts of the peroxide, and then titrating, using a molybdate catalyst under CO_2 to ensure complete reduction of the peroxide by HI. If insufficient HI is added, the iodine first liberated is bleached to HIO_3.

A sample of the crude aqueous peroxide, after repeated removal of water in the manner indicated above (about 80% of the whole), gave an average molecular weight of 81, and contained about one-quarter of its molecules as peroxides. Since distillation reveals considerable amounts of acetaldehyde, the molecular weight of the peroxide would be larger than this value. Reduction of this peroxide with zinc and glacial acetic acid gave rise to a complex mixture, which could not be worked up successfully.

The conclusions from these experiments was that the peroxide from the slow combustion of pentane is in part a vinyl ether peroxide from the peroxidation of dihydropyran, etc., whose formation has previously been substantiated. At present it is not possible to indicate whether this is the sole source.

Tests on the Substance X Responsible for the Absorption Band

The absorption band due to the formation of this substance was first detected at the end of the pseudo-induction period for hydrocarbons such as butane and pentane†; the work of Brunner‡ has shown that for such fuels the end of the pseudo-induction period coincides with the decomposition of the peroxides, etc., formed during a period of no pressure change. This suggests that X is formed by the decomposition of peroxides, which are thus an earlier product of combustion. Experiments with larger amounts of peroxide support this view.

---

* ' J. Amer. Chem. Soc.,' vol. 56, p. 2034 (1934).

† Egerton and Pidgeon, ' Proc. Roy. Soc.,' A, vol. 142, p. 26 (1933).

‡ ' Helv. chim. Acta.,' vol. 11, p. 881 (1928).

At first it was thought that the peroxide itself was responsible for the band, since the addition of ammoniacal lead hydroxide to the crude aqueous layer removes the band completely. This effect was actually due to the ready oxidation of X by the lead peroxide formed.

In spite of the high-absorption coefficient, which facilitated a study of the chemical behaviour, a final analysis has not yet been possible with the relatively minute amounts available from large quantities of pentane. A summary of its properties follows.

(*a*) Probably formed by decomposition of a peroxide, and distils in steam. A concentrate salted out with NaCl from the 95/98° distillate was used in further tests.

(*b*) Boiling-point $>$ 100°, but not very much above 100°. Melting-point $-16°$ C, fairly soluble in water.

(*c*) Probably not reactive with bromine in glacial acetic acid, and molecular weight not very different from 80. (With Br_2 only 10% of the total reacted.)

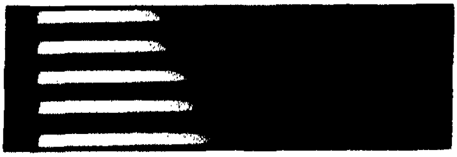
(*d*) Decomposes on standing in air, gradually oxidizing to a dark brown liquid with faint action on acid potassium iodide $F > M > B$. Decomposition probably accelerated by K_2CO_3.

(*e*) Readily oxidized by lead peroxide in ammoniacal solution, and $FeSO_4$ plus a peroxide in acid solution.

(*f*) High-absorption coefficient for ultra-violet over the region 2800–2500 A.

As has been stated in the preceding section, the absorption region shown by X resembles that of ketones, except for its position in the spectrum. The chemical properties, coupled with the ready change in the presence of acids and potassium carbonate, would suggest an activated carbonyl group as in keto-enol tautomers, but the reaction with phenyl hydrazine and 2·4 . phenyl-hydrazine was slow and possibly due to an impurity of dihydropyran, rather than a carbonyl group in the compound responsible for the band (*cf.* the action of Br_2).

In this connexion, the absorption spectra of ascorbic acid and its derivatives are perhaps significant in throwing light on the structure. In aqueous solution, ascorbic acid gives rise to a single very intense band with head at 2600–2650 A, which is thus practically coincident with the X band. A further similarity is that this band rapidly diminishes in

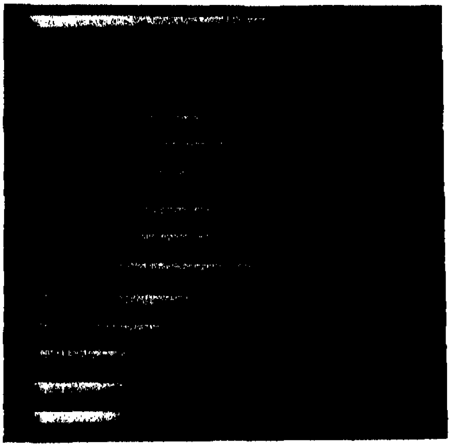


3

6

4

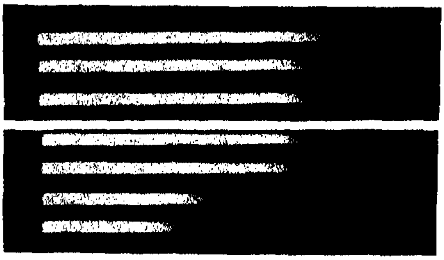
7



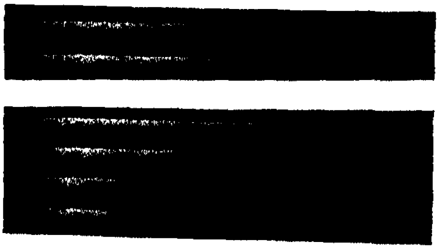
5

8

9

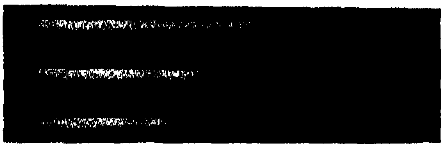


10



11

12



13

14

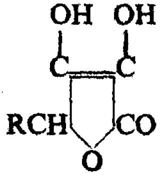
15

16

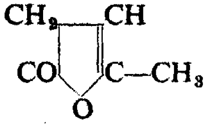
17

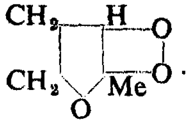
intensity (in a few hours) owing to air oxidation,* and that ascorbic acid is very readily oxidized with the usual reagents.

The structure



for ascorbic acid which is probably responsible for the intense absorption band may be compared with that of the angelilactones, which give a similar band on oxidation, and with that of the peroxides, which probably give rise to small quantities of X on destructive distillation.

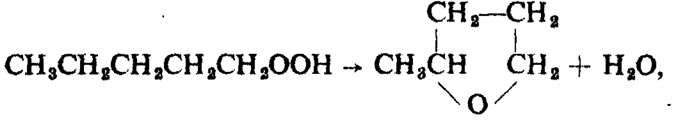




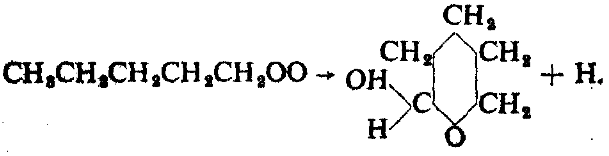
In spite of its intrinsic interest, further investigation of this substance is hampered by the very poor yields. Furthermore, it seems to be a secondary product of combustion, and probably does not influence the main reaction appreciably.

3—Conclusions

The conclusion from these investigations is that the formation of O ring compounds is of considerable importance in elucidating the oxidation mechanism of the higher paraffins at low temperatures. It is suggested (*cf.* Part I) that the formation of saturated O rings, such as methyl-tetrahydrofuran, corresponds with an internal oxidation by the energy-rich peroxide



and that the formation of unsaturated O rings occurs when a similar process takes place with an energy-rich peroxide radical



---

* Herbert, Hirst, Percival, Reynolds, and Smith, 'J. Chem. Soc.,' p. 1280 (1933).

Although these two compounds are the only ones whose presence has been definitely indicated, other modes of ring closure are not excluded, though the products are more difficult to identify on account of their labile structure in the presence of acids and less characteristic spectra.

The formation of peroxides which combine the properties of an olefine and an alcoxy peroxide is probably due to the peroxidation of unsaturated O ring compounds, since the vinyl ether peroxides which would be formed have the required behaviour.

The formation of a substance X with intense absorption band is probably due to the decomposition of a peroxide first formed. Its structure may resemble that of ascorbic acid.

The author wishes to thank Mr. A. C. Egerton, F.R.S., for much interest and help, and the Department of Scientific and Industrial Research foı a Senior Research Award.

APPENDIX

NOTES ON EXPERIMENTAL DETAILS

Purification of Ketones—This was most economically effected, with the small amounts available, by fractional freezing, using a special suction pipette. One-sixth and five-sixths of the ketone are frozen alternately with stirring, and the middle fraction was purified till it melted over a range of 1°, as determined with a calibrated thermocouple and sensitive galvanometer.

Preparations

Butylene—By dehydration of *n*-butyl alcohol with $H_2SO_4 + Al_2(SO_4)_3 + H_3PO_4$, washed with water, liquefied, and washed with conc. NaOH before use. Gave HCHO and faint CH_3CHO, but no X band on burning.

Pentane—From a crude cut boiling 30–42°, and practically free from unsaturated compounds. Distilled twice 34–38° with column, from conc. H_2SO_4, washed, and distilled 35–37° from solid KOH. *Iso-pentane* purified in a similar way, but specially fractionated 30·75–31·5° at 757 mm.

Hexane—Freed from cyclohexane, etc., by adding benzene and freezing out with CO_2, and removing uncrystallized benzene with sulphuric and nitric acids. Finally distilled over solid KOH. Boiling-point 67·25–67·3° short thermometer.

Furyl Alcohol—Redistilled 169–171·5° C.

Furfuraldehyde—Redistilled 160–161·5°, then purified by fractional freezing. When pure, furfuraldehyde was peroxidized by a stream of air or oxygen, though the peroxide did not accumulate in large amounts.

Furoic Acid—By oxidation of furfuraldehyde with cold alkaline $KMnO_4$, purified by sublimation. Melting-point 132° C.

Methyl Furan—(Boiling-point 62–64·5°) by decomposition of furfuraldehyde hydrazone with alcoholic sodium ethoxide. Salted out, and redistilled after drying over $CaCl_2$. Pure compound gives no reaction with KI, but gummed on standing, and gives strong peroxide tests. A small sample was specially purified by fractional freezing, but the boiling-point was not redetermined.

Tetrahydrofuryl Alcohol—Redistillation (175·5–176·5) of a sample obtained from the Quaker Oats Company. Optical tests showed practically no furyl alcohol.

Tetrahydrofuryl Bromide—By the action of PBr_3 on pure tetrahydrofuryl alcohol + pyridine, distilled twice *in vacuo*, fraction 68·5–70° per 16 mm collected. The Grignard reaction[*] was difficult to start in the cold, but it was found that the alcohol $CH_2 = CH . (CH_2)_3 OH$, which was the chief product (boiling-point 135–138° C), could be readily converted to *methyltetrahydrofuran* (boiling-point 79·5–80·5) by distillation with syrupy phosphoric acid, drying over K_2CO_3, and repeated distillation. The same substance was also prepared by hydrogenation of α-methylfuran with H_2 + Pt black, and had identical properties. $\mu^{17^\circ}_{\mathrm{NaD}}$ 1·406 F.

Tetrahydrofurylmethylether—By the action of methyl sulphate on pure tetrahydrofuryl alcohol in alkaline solution. Excess alkali was added to remove unchanged sulphate, and the product, after drying the ethereal extract over $CaCl_2$, gave a fraction boiling-point 144·5–146·0° C. $\mu^{17^\circ}_{\mathrm{NaD}} = 1{\cdot}437$.

Tetrahydrofuryl Ethyl Ether—Prepared in the same way, using ethyl sulphate. Boiling-point 158·5–160° C. $\mu^{17^\circ}_{\mathrm{NaD}}$ 1·438.

Furan—By decarboxylation of furoic acid, using a copper/quinoline catalyst.[†] Boiling-point 31·5–33·5°. Furan does not peroxidize as easily as α-methyl furan in air.

Acrolein from glycerol + syrupy phosphoric acid.[†] The crude product contained an impurity with absorption bands similar to α-methyl furan, but this disappeared on redistillation. Boiling-point 52–53° 756 mm.

---

[*] Paul, 'Bull. Soc. chim. Fr.,' vol. 53, p. 417; vol. 54, p. 1489 (1933).

[†] *Cf.* "Organic Syntheses."

Allyl alcohol (*ibid.*) from glycerol + formic acid purified by repeated distillation, final boiling-point 94–97°. Faint lines similar to dihydropyran in the absorption spectrum of the crude product disappeared on redistillation.

Dihydropyran—First obtained by the interaction of P_2O_5 on tetrahydrofuryl alcohol, later by dehydration over Al_2O_3 according to Paul* at 380° C. The product was dried over K_2CO_3, and distilled three times, finally collected 84·5–85·5° per 760 mm. The dihydropyran bands were still visible in the vapour when the concentration in hexane solution was 6%, and the cut-off at 0·4%, from which estimates of the amounts in products of slow combustion could be made. In aqueous solution these amounts are altered. *δ-oxyvaleraldehyde* (made from dihydropyran by catalytic addition of H_2O) gave no characteristic absorption with the vapour at 50° C.

Butyrolactone—(Boiling-point 202/205°) from the reduction of succinic anhydride in ethereal solution with sodium amalgam.

Levulinic acid from sucrose, according to "Organic Syntheses," *op. cit.*

α- and β-*Angelilactones* (*cf.* Vanino)—By very slow distillation of levulinic acid. The fraction boiling at 50/65° at 23 mm was purified by five fractional freezings in liquid air. The fraction 65/95° 23 mm was boiled with water for 2 hours, extracted with ether, dried over K_2CO_3, and redistilled 83–84°. Both these lactones gave a strong smell of burnt sugar and a band similar to X on slow combustion.

γ-*Valero lactone* by reduction of sodium levulinate with sodium amalgam (Vanino). Boiling-point 205·75–206·25° 754 mm. This lactone gave only a very faint X band on slow combustion, which later burnt away again.

Iso-propyl sulphate prepared according to Levaillant,† and purified by fractional freezing (melting-point −18° C).

Iso-propyl hydroperoxide according to Medwedew and Alexejewa.‡

Iso-butyl sulphate (*cf. iso*-propyl-sulphate) not so easy to purify owing to poor crystallization (melting-point −77/78°).

Iso-butyl hydroperoxide obtained in strong aqueous solution only, *cf. iso*-propyl hydroperoxide.

Dimethyl keten by the action of zinc on an ethereal solution of brom-*iso*-butyryl bromide, under CO_2. The solid peroxide is precipitated from the crude ethereal distillate by the action of air. The substance was introduced into the engine by spraying the ethereal solution into the throttle, and comparing with pure ether.

---

* 'Bull. Soc. chim. Fr.,' vol. 53, p. 417; vol. 54, p. 1489 (1933); vol. 1, pp. 971, 1397 (1934).

† 'C.R. Acad. Sci., Paris,' vol. 188, p. 261 (1932).

‡ 'Ber. deuts. chem. Ges.,' vol. 65, p. 133 (1932).

Alcoxy hydroperoxides—The various peroxides of the type RCH(OH) OOH and RCH(OH)OORCH(OH) were prepared according to methods described by Rieche* and used in the crude state for knock tests, etc.

Tetrahydropyran by hydrogenation of dihydropyran over platinum black. Boiling-point 86/90°† has penetrating ethereal smell. The absorption spectrum of the product was free from dihydropyran lines, and did not form dihydropyran on slow combustion.

Epoxy 1 . 4 pentene 5 according to Paul† from tetrahydrofuryl bromide + solid KOH, best redistilled in CO_2 owing to ready peroxidation. Boiling-point 81·5/83·5°, converted to *epoxy* 1 . 4 *pentene* 4 by dilute acids, with blackening (via *aceto*-propyl alcohol $CH_3CO(CH_2)_3OH$. Boiling-point 80/83·5°).

DESCRIPTION OF PLATES

PLATE 5

FIG. 3—X vapour from crude aqueous condensate from hexane, in oxygen. Temperature gradually raised from 20°–180°.

FIG. 4—Vapour of X from pentane combustion at 20°, as salted out with NaCl from distillates. Length of column of vapour in nitrogen gradually decreased by adding pure nitrogen.

FIG. 5—Gradual heating of hexane/air NO_2 from 18°–320° C. The first exposure gives the length of the available spectrum in pure air. The gradual disappearance of NO_2 even before the band appears may be due to thermal dissociation and to interaction with the hexane.

FIG. 6—Gradual combustion of furyl alcohol vapour in oxygen, 150°–400° C. (Vapourizer at 0° C.)

FIG. 7—Slow combustion of tetrahydrofuryl ethyl ether (18·5° C) 200°–260° C, with formation of X-like band. (Vapourizer at 18·5° C.)

FIG. 8—Slow combustion of β-angeli-lactone at 240° (steady) in air. Product pushed along with N_2 in last three exposures. (Vapourizer at 19° C.)

PLATE 6

FIG. 9—Slow combustion of fraction (75/81°) from combustion of pentane after distillation over metallic Na (traces of dihydropyran still present are visible in the first few exposures). Vapourizer at 18° C, combustion 20–300° C.

FIG. 10—Slow combustion tetrahydropyran 20°–390° C, in air (vapourizer at 18° C).

FIG. 11—Slow combustion tetrahydromethyl furan, 1st and 2nd exposure after prolonged heating at 280° C, then fresh charge 39°–360° C (vapourizer at 18·5° C).

---

* "Peroxide and Ozonide," Dresden, 1931.

† 'Bull. Soc. chim. Fr.,' vol. 53, p. 417; vol. 54, p.1489 (1933); vol. 1, pp. 971, 1397 (1934).

FIG. 12—Furan vapour from 0°–160° C in air (vapourizer at 20° C).
FIG. 13—α-Methylfuran vapour (specially purified by freezing), vapourizer at 5°, 21°, 60° C.
FIG. 14—Dihydropyran and hexane vapour at 20° C, (i) pure hexane vapour, 50%; (2), 25%; (3), 12%; (4), 5%; and (5), 3% of dihydropyran.
FIG. 15—Methyl ethyl ketone and acetone, vapourizer at 0°.
FIG. 16—Amyl nitrite in N_2 from 20°–320° C, vapourizer at 17° C.
FIG. 17—Purified ethyl ether, slow combustion in air from 115°–200° C.

The Absorption of Monochromatic X-Ray Beams, of Wave-length in the Region 50 to 20 *X*-Units, in Lead, Tin, Copper, and Iron

By JOHN READ, Ph.D., Kellogg Radiation Laboratory, California Institute of Technology, Pasadena

(*Communicated by J. C. McLennan, F.R.S.—Received May* 27, 1935)

[PLATE 7]

INTRODUCTION

In a previous paper an account has been given of the measurement of the absorption of monochromatic X-ray beams of wave-length in the region 50 to 20 *x*-units, in carbon and aluminium.* The relation of the measured coefficient of absorption to the wave-length did not differ from that predicted by the Klein-Nishina formula by more than 1%. The method used in that experiment has been improved, and used to measure the absorption coefficients of lead, tin, copper, and iron for similar monochromatic beams. Because lead has been used very extensively for absorption measurements the primary aim has been to measure as accurately as possible the dependence of its absorption coefficient on the wave-length of the radiation. It has not been possible to make such accurate measurements on tin, copper, and iron, but enough data has been obtained to determine the variation of the photoelectric absorption coefficient per electron with the atomic number of the absorbing element, with fair accuracy, for radiation in this region of wave-lengths.

* Read and Lauritsen, 'Phys. Rev.,' vol. 45, p. 433 (1934).

Since these absorption coefficients may find considerable application, it is considered well to give a more detailed account of the method of measurement, so that an independent judgment of their reliability may be made.

APPARATUS

The arrangement of the apparatus is shown in fig. 1. The source of radiation was an X-ray tube A, which could be excited by voltages up to 1000 kV. A spectrograph was used to select radiation of a desired wave-length from the continuous spectrum. Two lead blocks B and C had slits, each 0·02 in by 1 in wide by 3 in deep, which defined a plane horizontal sheet of radiation. This passed through a rock-salt crystal D, incident on its internal atom planes at such an angle that the desired wave-length was reflected in a vertical plane. The lead block E, having a slit of vertical width 0·04 in, allowed the monochromatic beam to pass through but absorbed the unreflected radiation. The beam then passed into the next room through an aperture 12 in square in a concrete wall 10 in thick. Stray radiation was excluded from this room by lead plates G 2 in thick, and a lead block H 6 in long pierced by a 0·25-in slit to pass the beam. A lead filter F was chosen of a suitable thickness to cut down the soft scattered radiation without weakening too seriously the monochromatic beam. The latter could be shut off by a sliding lead block J, or allowed to pass through two ionization chambers K and L. Each chamber was an aluminium box 8 by 8 by 4 in, shielded from stray radiation by lead blocks 3 in thick at the front and back, $\frac{5}{8}$ in at the sides, and $\frac{3}{8}$ in at the top and bottom. The lead blocks at the front and back had slits to pass the beam, and all slits except B and C were made just so large that the beam did not graze their sides. Internal partitions limited the effective width of the ionization chambers to a little greater than the width of the X-ray beam.

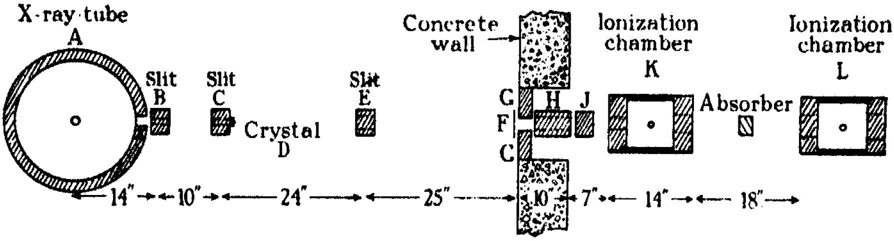


FIG. 1—Arrangement of the apparatus.

The construction of the ionization chambers is shown in fig. 2. A low-power microscope A projected into the aluminium box. The brass block B carried an amber rod C 1 cm long and 1 mm in diameter, which insulated the wire frame D. To this frame was attached a gold-coated L-shaped quartz fibre E, about 2 μ in diameter. It was 7 mm long and had a " foot " 0·5 mm long. The " foot " could be brought into focus in the microscope by adjusting the block B, and when charged its image moved across an eye-piece scale. It was charged by rotating the wire F to touch the wire frame D. When not in use the wire F was earthed in a fixed position. Maximum deflection of the fibre on the scale was obtained with a potential difference of about 400 volts between it and the box.

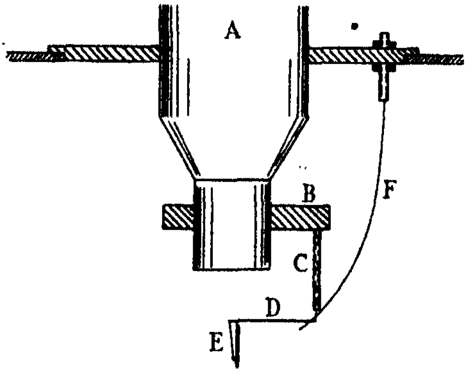


FIG. 2—Construction of the ionization chamber.

EXPERIMENTAL PROCEDURE

For each chamber it was necessary to know the relation between a discharge through any range on the eye-piece scale and the quantity of radiation producing it. Since the experiment required comparison of the intensities of beams of the same wave-length, ionization could be used as a measure of intensity. Some radium was placed in line with the slit system of the chambers, and photographs were taken at equal time intervals of the positions of the fibres as they discharged. Graphs with scale reading as ordinate and time as abscissa were then constructed. Since the loss of charge due to insulation leakage was negligible compared with that due to the gamma rays and cosmic rays, equal times of discharge were equivalent to equal quantities of ionization. Therefore a unit on the time axis could be taken as an arbitrary unit of ionization, and the ionization causing a discharge through any range on the eye-piece scale could be expressed in terms of this unit. To determine whether the response of the chambers was affected by the size of the ionization current the radium was placed at different distances from the chambers so that discharge rates of 5, 10, and 15 divisions per minute were obtained. It was found that discharge rates of less than 10 divisions per minute gave graphs of the same shape when the time unit was suitably chosen. In the experiment the discharge rates were therefore kept below 10 divisions per minute.

It was also found that the ionization produced by two radioactive sources acting together equalled the sum of the ionizations produced by the sources acting separately.

MEASUREMENT OF ABSORPTION

When the two fibres had been charged, shutter J was opened for a time chosen so that the more rapidly moving fibre covered almost the whole eye-piece scale. This was done with no absorber between the chambers. The ionization causing each discharge could be determined in arbitrary units from the radium discharge curves, so that the ionization in chamber L could be expressed in terms of that in chamber K. The discharge was repeated with an absorbing screen midway between the chambers. From the ionization in chamber K, and the known response of L in terms of K when no absorber was present, the ionization which would have occurred in chamber L had no absorber been present was calculated. The actual ionization in L was known from the observed discharge, so that the absorption coefficient could be found. The ionizations were always corrected for ionization due to general and cosmic radiation. The correction was found by altering the angle of the crystal so that it no longer reflected the monochromatic beam down the slit system. The discharge of each chamber was then measured under the same conditions as existed during the measurements to be corrected. To minimize the effect of any change in the ratio of the response of chamber L to that of K, this ratio was determined before and after measurements with an absorber, and the mean was used.

The absorption coefficient was measured for several different thicknesses of absorber to see whether there was any hardening of the beam with increasing thickness. In almost every case it was evident that this did not occur.

The X-ray tube was operated at a peak voltage less than twice that corresponding to the wave-length reflected by the crystal. Therefore radiation of half the desired wave-length, reflected in the second order, was not present.

It can be shown that an absorber thickness which reduces the intensity of the beam by about 70% minimizes the final error due to errors of scale reading. The absorber thicknesses were chosen so as to cut down the intensity of the beam by from 50% to 80%.

A sample set of observations is given in Table I. k_1 and k_2 are constants of proportionality between the arbitrary units of ionization and a true unit. When the absorber was present the ionization in chamber K was 1653 k_1. Hence, had no absorption occurred, the ionization in chamber L would have been $1653\, k_1 \times 1 \cdot 2459 \frac{k_2}{k_1} = 2059\, k_2$. Actually it was 1422 k_2, so that $I_0/I = 1 \cdot 4480$ and $\mu t = \log 1 \cdot 448$, where I_0, I, μ, and t are related by $I = I_0\, e^{-\mu t}$.

MEASUREMENT OF THE WAVE-LENGTH AND ITS RESOLUTION

To determine the wave-length of the radiation admitted to chamber K a photograph of the direct and reflected beams was taken at a distance of 167 cm from the crystal. Here the separation of the lines was from 1 to 3 cm. This separation was measured by taking microphotometer traces perpendicularly across the lines. It was possible to obtain only one photograph for each wave-length, but two traces were taken across different parts of each. In the majority of cases the two values obtained for the wave-length from each pair of traces differed by less than 2%. It is unlikely that the peak wave-lengths differed by more than ± 2% from the calculated values. Fig. 3, Plate 7, is a reproduction of a photograph showing the direct and reflected beams.

To determine the wave-length resolution ionization chamber K was moved in steps normally across the reflected beam. The ionization in each position was noted and is plotted in fig. 4. A wedge of radiation from a true line source which just filled slit C, fig. 1, would, on reflection from a true crystal, give a band of reflected wave-lengths of width 4·5 x-units. Since the slit of chamber K had a width equivalent to 11 x-units this band would appear to have a width of 15·5 x-units, when explored in the manner described above. The diagram shows a width of about 16 x-units, so that it would seem that the focal spot did indeed act very nearly as a line source giving, on reflection, a band of wave-lengths of width about 5 x-units. Chamber K was then fixed and the crystal was rotated in small steps, the ionization in chamber K for each position of the crystal being noted. The angle of rotation of the crystal within which ionization occurred corresponded to a band of wave-lengths of width 10 x-units. This also is compatible with reflection of a band of

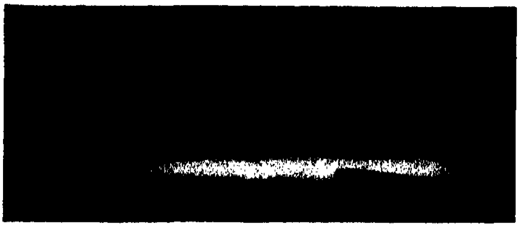


FIG. 3—Photograph of direct and reflected beams at a distance of 167 cm from the crystal.

TABLE I—SPECIMEN SET OF OBSERVATIONS

Experimental arrangement	Time of discharge min	Chamber	Initial scale reading	Final scale reading	Initial reading from radium curve	Final reading from radium curve	Ionization of monochromatic beam plus background	Ionization without monochromatic beam	Background ionization	Ionization due to monochromatic beam alone	Ratio of ionization in L to that in K	Mean ratio
No absorber present	10	K	90·3	27·15	303	1755	1452 k_1	11·4 k_1	114 k_1	1338 k_1	1·2399 k_2/k_1	—
		L	94·3	8·15	195	2032	1837 k_2	17·8 k_2	178 k_2	1659 k_2		
Absorber in position	12	K	89·9	13·0	312	2102	1790 k_1	11·4 k_1	137 k_1	1653 k_1	—	1·2459 k_2/k_1
		L	98·3	20·75	113	1737	1624 k_2	16·8 k_2	202 k_2	1422 k_2		
No absorber present	10½	K	90·8	26·7	292	1766	1474 k_1	11·4 k_1	120 k_1	1354 k_1	1·2518 k_2/k_1	—
		L	93·0	5·15	222	2104	1882 k_2	17·8 k_2	187 k_2	1695 k_2		

true width (allowing for the width of the slit admitting radiation to chamber K) of about 5 x-units. Therefore, with a fixed position of the crystal and chamber K only a certain band of wave-lengths of this width could enter K, and movement of the focal spot (the only thing likely to move) would not change these wave-lengths but only move the intensity peak by ± 2·5 x-units.

SOURCES OF ERROR

Scattering of X-rays into Chamber L by the Absorber—X-rays scattered out of the beam, yet still able to enter chamber L, will cause the measured value of the absorption coefficient to be too small. This effect, per electron, will be less than that due to that particular electron in the absorber which is most efficient in scattering into chamber L. This electron is on the axis of the slit admitting radiation to L, and nearest to the slit. The slit was 4 in by ¼ in, and its more distant face was 10¼ in from the nearest part of an absorber. By means of the Klein-Nishina formula the fraction of energy scattered by this electron through the slit can be estimated to be less than 10^{-27}, and therefore this effect can be neglected. Radiation due to the return of ejected photoelectrons, which enters chamber L, will also be negligible compared with the photoelectric absorption, since only a fraction of the latter reappears as radiation and it is emitted isotropically.

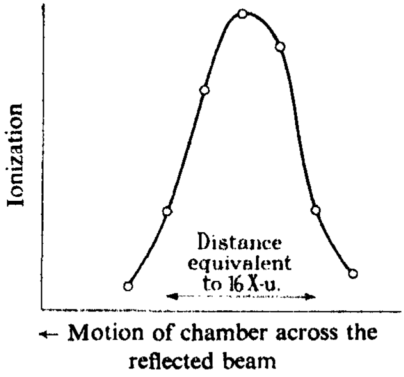


FIG. 4—Energy distribution of the reflected beam as explored by chamber K.

Lack of Homogeneity of the Beam—Since the reflected beam has a wave-length spread of about 6 x-units the measured absorption coefficient will not correspond to the peak wave-length. It can easily be shown that if μ_1 cm^{-1} represents the true absorption coefficient corresponding to the peak wave-length, and if μ_2 cm^{-1} represents the measured absorption coefficient, if x represents the half-width of the line and d the thickness of the absorber in centimetres, then, assuming a shape for the line illustrated by fig. 4,

$$\mu_1 - \mu_2 = \frac{x^2 d}{8}\left(\frac{d\mu}{d\lambda}\right)^2.$$

This correction is negligible except for lead, when it is of the same order of magnitude as the likely experimental error.

RESULTS

Lead—The results for lead are set out in Table II. Column 2 gives the difference between the two values obtained for each wave-length as a percentage of the mean. Column 10 gives the difference between the maximum and minimum values of the absorption coefficient obtained for each wave-length, expressed as a percentage of the mean. The values, per electron, are graphed in fig. 5. The height of each black rectangle is sufficient to include the sets of values for each wave-length listed in column 8. The width of each rectangle is sufficient to include the two measurements of the wave-length, with the addition of 1% to cover other possible errors, such as error in measuring the distance from the crystal to the photographic film. It is very unlikely that the true values lie outside of these rectangles. A smooth curve (1) can be drawn through all of them, and this curve cannot be varied by as much as 2% at any point without missing some, or introducing sudden changes of slope which are unlikely in this region of wave-lengths. It is considered that this curve gives the electronic absorption coefficient for a given wave-length, or the wave-length corresponding to a known absorption coefficient, with an error of less than 2%. Curve 4 gives the Klein and Nishina scattering per electron. The differences between the ordinates of curve 1 and curve 4 are plotted to form curve 2. This difference is usually ascribed to photoelectric absorption. The empirical photoelectric law of L. H. Gray* is illustrated by the dotted curve 3. Gray's curve is some 20% lower at 20 x-units and 7% lower at 50 x-units wave-length than the experimental curve. The measured values for the absorption coefficients for lead are given in Table III.

Values for the photoelectric absorption coefficient of lead for longer wave-lengths have been published by Allen.† His value of 41·7 cm^{-1} at

---

* 'Proc. Camb. Phil. Soc.,' vol. 27, p. 103 (1931).

† 'Phys. Rev.,' vol. 27, p. 266 (1926).

TABLE II—EXPERIMENTAL RESULTS FOR LEAD

1 Mean wave-length λ in *x*-units	2 $\frac{\Delta\lambda}{\lambda} \times 100$	3 Screen thickness mm	4 Number of values obtained	5 Mean μ cm^{-1} for each screen	6 Probable error as %	7 Correction for inhomo-geneity	8 Corrected μ cm^{-1}	9 Mean μ cm^{-1}	10 $\frac{\Delta\mu}{\mu} \times 100$	11 $\mu_e \times 10^{25}$
52·5	1·0	1	4	7·687	0·5	0·017	7·704	7·75	1·7	29·3
		2	4	7·756	0·3	0·034	7·790			
48·5	2·9	1	2	6·756	0·5	0·009	6·765	6·605	4·7	25·0
		1½	2	6·607	0·7	0·014	6·621			
		2	4	6·411	0·7	0·018	6·429			
41·6	1·2	1½	2	4·358	0·6	0·007	4·365	4·41	3·2	16·7
		2	2	4·472	0·4	0·010	4·482			
		2½	2	4·320	0·3	0·012	4·332			
		3	2	4·445	0·9	0·014	4·459			
38·4	0·5	2	4	3·838	0·5	0·007	3·845	3·89	2·8	14·75
		3	4	3·933	0·5	0·010	3·943			
31·1	1·2	3	4	2·562	0·7	0·005	2·567	2·54	1·6	9·63
		5	4	2·509	0·3	0·008	2·517			

28·4	4·2	3	4	2·127	0·8	0·003	2·130		3·3	8·06
		5	4	2·091	0·2	0·005	2·096	2·13		
		7	4	2·153	0·4	0·007	2·160			
25·2	2·7	5	4	1·910	0·8	0·006	1·916	1·95	3·4	7·37
		7	4	1·968	0·5	0·008	1·976			
24·8	3·6	5	4	1·930	0·2	0·006	1·936	1·94	—	7·33
22·8	1·3	5	4	1·687	0·7	0·005	1·692		1·4	6·43
		6	4	1·706	0·4	0·006	1·712	1·70		
		7	4	1·681	0·5	0·007	1·688			
21·6	3·7	5	4	1·556	0·6	0·004	1·560	1·51	6·6	5·71
		7	4	1·449	0·6	0·006	1·455			
20·2	1·5	5	4	1·503	0·7	0·004	1·507	1·50	0·5	5·68
		6	4	1·492	1·0	0·005	1·497			
		7	4	1·496	0·6	0·006	1·502			

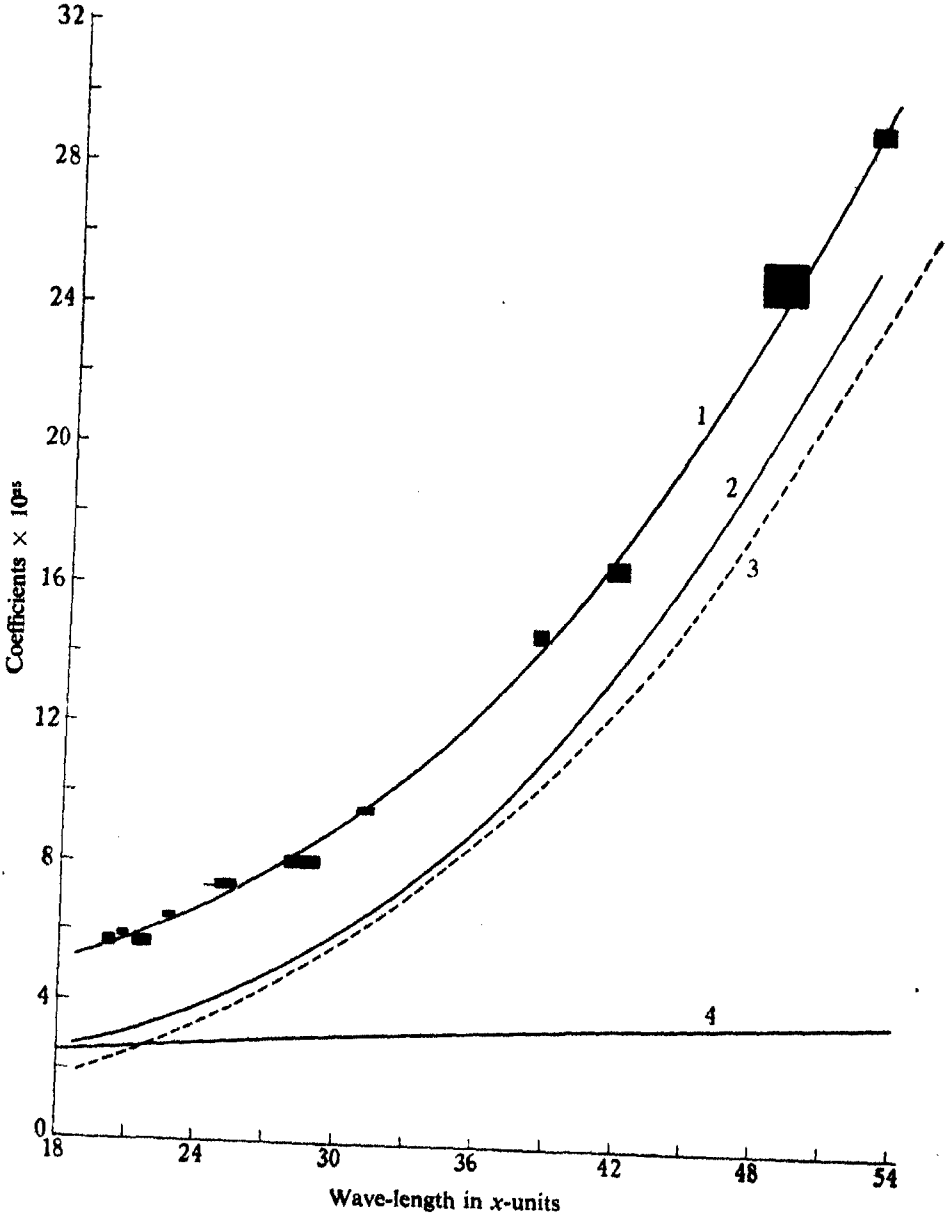


Fig. 5—(1) Experimental values of μ_e for lead; (2) experimental values of τ_e for lead; (3) Gray's empirical photoelectric law; (4) Klein and Nishina values for σ_e.

102 x-units and the values listed above may be linked up by the empirical formula:

$$\tau \text{ cm}^{-1} = 0{\cdot}01377\ \lambda + 0{\cdot}0003326\ \lambda^2 + 0{\cdot}00003474\ \lambda^3,$$

where λ is expressed in x-units.

Tin, Copper, and Iron—The values obtained for τ_e for tin, copper, and iron are given in Table IV. For each wave-length it was not possible to make as many observations as with lead, but the values are probably not in error by as much as 5%.

TABLE III—ABSORPTION COEFFICIENT FOR LEAD

λ in x-units	$\mu_e \times 10^{25}$	$\sigma_e \times 10^{25}$	$\tau_e \times 10^{25}$	μ cm^{-1}	τ cm^{-1}
21	5·80	2·67	3·13	1·53	0·827
24	6·64	2·82	3·82	1·75	1·01
27	7·80	2·97	4·83	2·06	1·28
30	9·16	3·10	6·06	2·42	1·60
33	10·66	3·21	7·45	2·82	1·97
36	12·50	3·32	9·18	3·30	2·42
39	14·80	3·42	11·4	3·92	3·00
42	17·40	3·52	13·9	4·60	3·67
45	20·36	3·61	16·7	5·37	4·43
48	23·70	3·70	20·0	6·26	5·29
51	27·40	3·77	23·6	7·25	6·25

In the region of longer X-ray wave-lengths it has been found that the electronic photoelectric absorption coefficient τ_e varies as the cube of the atomic number Z of the absorbing element. When the values for τ_e for lead, given by curve 2 of fig. 5, and this Z^3 law are used to calculate τ_e for tin, copper, and iron, the curves shown in fig. 6 are obtained. If the values of the Klein and Nishina absorption coefficient τ_e are then added to the values of τ_e, three curves are obtained which can be compared with the experimental points. For copper and iron the agreement is good, the deviations of the points from the curves being less than the likely experimental error. However, the deviations for tin are much larger than can be accounted for in this way, and they fit much better the curve which has been obtained on the assumption that τ_e varies as $Z^{3.5}$. It would therefore seem that the variation of τ_e with Z cannot be described by a simple power law, and further experiment on other elements will be necessary to determine its manner of variation.

TABLE IV—RESULTS FOR IRON, COPPER, AND TIN

Element	Wave-length in x-units	Number of measurements of μ_e	Mean $\mu_e \times 10^{25}$
Iron	38·4	3	3·72
	41·6	4	3·99
	48·5	2	4·40
Copper	22·8	4	3·05
	23·2	5	2·98
	28·4	4	3·16
	38·4	3	3·77
	41·6	4	4·20
	48·5	2	4·80
Tin	22·8	4	3·68
	28·4	4	3·89
	43·8	5	6·33

I thank Dr. C. C. Lauritsen for suggesting this work and for his advice during its progress. I am also grateful to Dr. H. Dingle for undertaking part of the microphotometry. The work was made possible by a grant from the Seeley W. Mudd Fund, and thanks is offered for this assistance.

[*Note added in proof August* 8, 1935—It is of interest to compare the results of this experiment with the theoretical values recently published by Hulme, McDougall, Buckingham, and Fowler.* These authors have calculated the photoelectric absorption of a simplified atom model consisting of a fixed nucleus of atomic number Z, and one K electron, using the relativistic wave functions of Dirac. Their results should apply in the range not previously examined—where Z is large and $h\nu$, the quantum energy of the incident radiation, is of the order of magnitude of mc^2, the rest-mass energy of the electron. They have correlated their results with those of previous authors, who have examined other regions of Z and $h\nu$, and have published a set of curves giving the photoelectric absorption per atom for wave-lengths between zero and 36 x-units, and for a number of elements up to $Z = 82$. Their method of converting the calculated values for the atom model to values relating to a real atom was to double the calculated value to give the absorption due to the K shell, and to multiply this by 5/4 so as to include approximately the effect of the outer shells.

* 'Proc. Roy. Soc.,' A, vol. 149, p. 131 (1935).

The curves of fig. 7, which from top to bottom of the diagram relate to lead, tin, copper, and iron, in that sequence, have been obtained by adding the Klein and Nishina absorption per electron to the values for the photoelectric absorption per electron taken from the paper of Hulme,

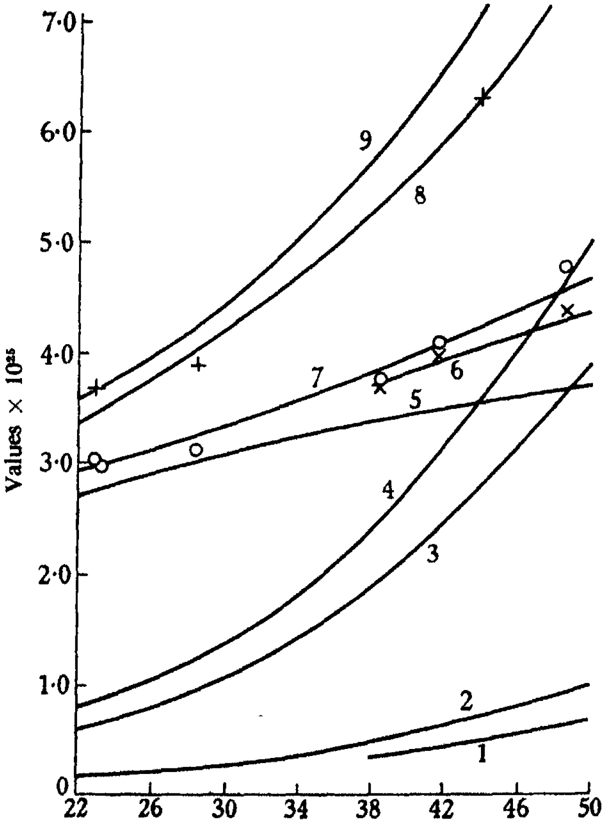


FIG. 6—(1) τ_e for iron calculated from $\tau_e \propto Z^3$; (2) τ_e for copper calculated from $\tau_e \propto Z^3$; (3) τ_e for tin calculated from $\tau_e \propto Z^{3\cdot5}$; (4) τ_e for tin calculated from $\tau_e \propto Z^{3\cdot0}$; (5) Klein and Nishina absorption σ_e; (6) μ_e for iron calculated on the assumption $\tau_e \propto Z^3$; (7) μ_e for copper calculated on the assumption $\tau_e \propto Z^3$; (8) μ_e for tin calculated on the assumption $\tau_e \propto Z^{3\cdot5}$; (9) μ_e for tin calculated on the assumption $\tau_e \propto Z^{3\cdot0}$. × experimental points for iron; ○ experimental points for copper; + experimental points for tin.

McDougall, Buckingham, and Fowler. The values for lead, tin, and iron were obtained directly from their paper, while the values for copper were obtained by interpolation. The centres of the circles of fig. 7 give the experimental values of this paper. Those for tin, copper, and iron *are the actual measured values, while those for the lead are taken from the* smoothed curve of fig. 4. The size of the circles indicates the region within which the true values probably lie. The theoretical curves for tin, copper, and iron agree quite satisfactorily with the experimental values (though this does not check the photoelectric absorption very closely since it is a small part of the whole). It would appear, however, that in the case of lead, for wave-lengths in the region 20 to 30 x-units, the absorption is greater than that calculated, and if the difference is all allotted to photoelectric absorption an increase of 25% in this coefficient would be necessary at 20 x-units. However, that the disagreement should exist for the shorter rather than for the longer wave-lengths makes its explanation difficult, since the calculations should apply more rigorously for the shorter wave-lengths. The experimental measurements became considerably more difficult in this region, and perhaps there was a systematic tendency for too high values to be obtained (compare the results for carbon obtained by the same method, and published in an earlier paper*). There is no reason to expect such a tendency, however,

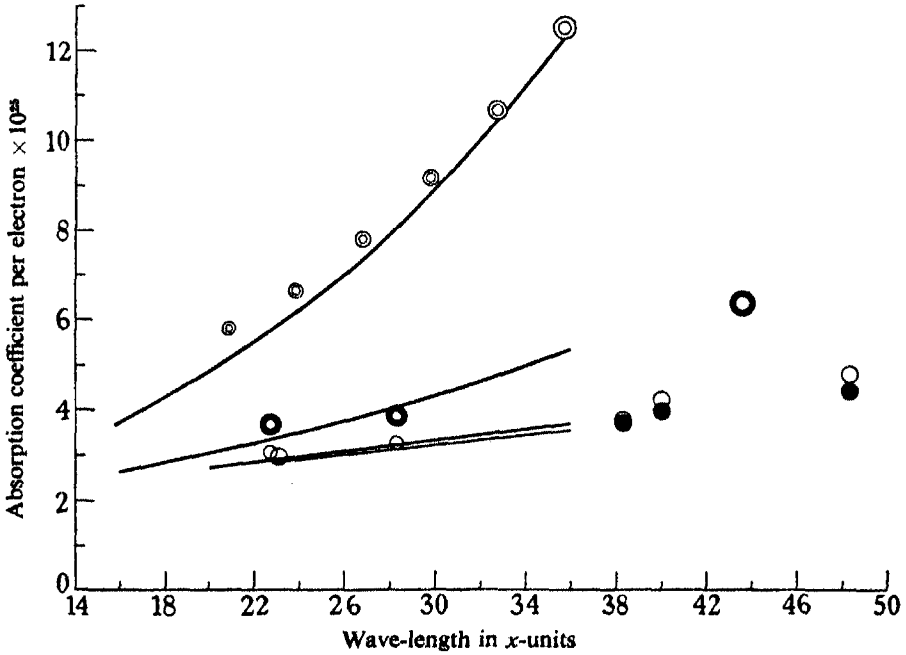


Fig. 7—Comparison of the experimental absorption coefficients with those calculated from the Klein and Nishina formula and the results of Hulme, McDougall, Buckingham, and Fowler. ◎ lead; O tin; ○ copper; ● iron.

* Read and Lauritsen, 'Phys. Rev.,' vol. 45, p. 433 (1934).

since the X-ray tube was always operated at a voltage almost twice that corresponding to the wave-length of the radiation under examination, and the beam was well filtered (1 mm of lead and 10 mm of steel, approximately) before reaching the spectrograph. There would, therefore, be as much radiation harder than that under examination as softer, and as much chance of the reflected beam being adulterated with harder as with softer radiation.

From a practical point of view, it is interesting to note that from the Klein and Nishina formula and the curves of Hulme, McDougall, Buckingham, and Fowler it would appear possible to calculate absorption coefficients for any element and any wave-length less than 36 *x*-units, with an error certainly not greater than 10% (provided electron pair production is not appreciable).]

SUMMARY

An experiment is described by which the absorption coefficients of monochromatic X-ray beams in the wave-length region 50 to 20 *x*-units in lead, tin, copper, and iron have been measured. A curve relating the absorption coefficient of lead and the wave-length, accurate to 2%, has been obtained. From it the photoelectric absorption coefficients have been calculated, and they are from 20 to 7% greater than corresponding values obtained from the empirical law of L. H. Gray. They may be linked up with Allen's value of 41·7 cm^{-1} at 102 *x*-units by the law:

$$\tau \text{ cm}^{-1} = 0\cdot01377\,\lambda + 0\cdot0003326\,\lambda^2 + 0\cdot00003474\,\lambda^3,$$

where λ is expressed in *x*-units.

It has been shown that the law $\tau_e \propto Z^3$ holds very closely for lead, copper, and iron, but lead and tin are related more nearly by $\tau_e \propto Z^{3\cdot5}$.

Explosion Waves and Shock Waves
III—The Initiation of Detonation in Mixtures of Ethylene and Oxygen and of Carbon Monoxide and Oxygen

By W. PAYMAN, D.Sc., Ph.D., and H. TITMAN, B.Sc.

(*Communicated by J. F. Thorpe, F.R.S.—Received May* 28, 1935)

[PLATES 8-14]

INTRODUCTION

This series of papers has so far dealt mainly with non-maintained or partially maintained atmospheric shock waves, and only incidentally with the fully maintained " detonation " wave. It is generally accepted that the detonation wave in an explosive gas mixture is a shock wave produced by the rapid combustion of the mixture, sufficiently intense to cause almost instantaneous ignition of the gas through which it passes, and continuously maintained by the combustion thereby started. An account of some preliminary experiments, using the " wave-speed " camera to record the movements of the flame and of the invisible shock waves in front of the flame in gas mixtures prior to detonation, has already been given by one of us*. Those experiments related mainly to hydrogen-oxygen and methane-oxygen mixtures whose aptitude to detonate may be regarded as moderate. For the continuation of the work, mixtures with oxygen have again been used, but a more readily detonating gas, ethylene, was chosen. Experiments were also made with carbon monoxide, because the flame usually requires a comparatively long run before detonation is established. These two gases have the advantage, not shared by hydrogen and methane, that their predetonation flames are sufficiently actinic for good records to be obtained by direct photography for comparison with corresponding " wave-speed " records. All gas mixtures used were saturated with water vapour.

THE EXPLOSION TUBE

The design of explosion tube to be used for the experiments was important. Fig. 1 illustrates it and a photograph of the component parts is given in fig. 2, Plate 8. That section of the tube containing the windows was constructed from a bar of stainless steel of rectangular cross-section, 5 cm by 6·5 cm, and 37 cm long. A hole 2·5 cm in diameter was drilled on an accurately centred line in this block, so forming the explosion tube. The metal on each of the wider faces of the steel bar was milled away, except towards each end, until the explosion tube was encountered; the final milling was then carried out carefully so that sharply edged slits, 4 mm wide and 30 cm long, were formed along opposite sides of the tube. A side view of the resulting tube is shown at *V* in fig. 1. Two window frames were constructed of exactly the same shape and size as the portions milled away, so that with these frames in position, as at *W* in fig. 1, the bar was once more of rectangular cross-section. Bevelled slots were cut in the frames to coincide internally with the slits in the tube. These slots were not made continuous, but were divided into equal lengths by two bars which served as ties, and as reference marks on the photographic records. The windows consisted of flat strips of cellophane, a strip 2 cm wide (thickness 0·2 mm) being placed over each slot and secured along its sides by adhesive tape. The window frames were then secured in position, after which the ends of the cellophane strips were also sealed with adhesive tape. When closed at both ends, the explosion tube held a vacuum, and the windows were sufficiently strong to withstand the explosion, and occasionally even the detonation, of gaseous mixtures. This part of the explosion tube will be referred to in future descriptions as the "steel section."

---

* Payman, 'Proc. Roy. Soc.,' A, vol. 120, p. 90 (1928).

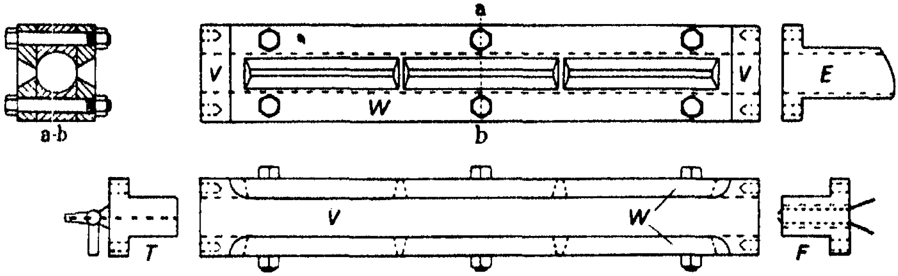


FIG. 1.

Special brass end-pieces were made, the one at *T* carrying a tap for evacuating and filling the vessel, the one at *F* insulated wires for use as a spark gap. When the igniting spark was to be passed at a point close to the end of the tube, two wires were used, as shown in fig. 1, the gap being then 3 mm from the end. When the spark was desired some distance away, the end-piece used carried a single insulated wire of suitable length placed away from the axis of the tube so as not to be visible through the windows; the end of this wire was pointed and bent over so as to terminate close to the wall of the tube, and the spark passed from the point to the wall. Most of the end-pieces were so constructed that the inner surface, forming the end of the explosion tube, coincided with one end of the windows. The ends were secured by studs and were made gas-tight by machining the faces of the flanges which were greased before the end-pieces were placed in position. In addition, the steel section could be attached at either end to flanged extension tubes of brass similar to the one shown at *E*. Pairs of dowel pins, not shown in fig. 1, provided for the accurate alignment of the various units.

INTERPRETATION OF MOVING-FILM RECORDS

The first attempt to deduce the behaviour of shock or sound waves in exploding gas mixtures from direct photographs appears to have been that made by von Oettingen and von Gernet.* Their interpretation of the records obtained, using a stationary film and a revolving mirror, was shown later by Dixon† to be incorrect, and he developed a system of deduction which he applied to his own records, obtained with a moving film. In his system the speed of a shock wave was usually assumed to be equal to that of sound in the gas mixture, and its history was deduced from its visible effects on the flame. Valuable as direct photographic records are for depicting the wave movements through the hot and luminous gases, they give but little information about the passage of the waves through the unburnt gas, their speeds and, often, their points of origin being unknown.

The wave-speed camera affords a means of obtaining the missing information. For a general explanation of the type of Schlieren records obtained reference should be made to a previous paper.‡ It may be pointed out that in all the records reproduced in the present paper the initial flame travels from right to left and the film travels upwards. Thus the distance scale is horizontal, reading from the right, and the time scale is vertical, reading downwards. The distance scale is constant, except in figs. 35 and 36, representing a length of 30 cm, but the time scale (speed of movement of the film) has been varied according to the speed of the flame. The distance between successive marks placed to the right of each record represents a time interval of 1 millisecond. The left-hand side of each record usually corresponds with one closed end of the explosion tube. The records are reproduced in this paper at approximately five-sixths their original dimensions.

---

* 'Ann. phys. Chem.,' vol. 33, p. 586 (1888).

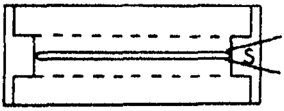
† 'Phil. Trans.,' A, vol. 200, p. 315 (1903).

‡ Payman, *loc. cit.*

The line diagrams which follow all refer to the explosion tube and are representations, not to scale, of the conditions for each experiment or set of experiments. In each diagram S indicates the point of ignition and R and E the positions of restrictions or enlargements of the cross-section of the tube. The records to which a line figure refers are denoted in corresponding Roman numerals beneath it.

EXPERIMENTAL

Flame and Shock Waves in a Comparatively Slowly Burning Mixture—Figs. 3 to 5, Plate 8. For the first series of experiments, the steel section was used alone as the explosion tube. The length of the tube when closed at both ends was therefore 30 cm, and the whole progress of the flame was photographed. A rather dilute mixture of ethylene and oxygen, containing 15% of ethylene, was used and was ignited close to one end of the tube by an electric spark of moderate intensity—a break spark from a small induction coil. Wide variations could be made in the energy of this spark, by alteration in the primary current passing through the induction coil, without any apparent change in the mode of propagation of the flame.



III, IV, V.
XXIV, XXV, XXVI, XXVII, XXVIII, XXIX.

Fig. 3 is a reproduction of a moving-film record, obtained by direct photography, of the passage of flame through the mixture. The spark can be seen as a white spot near the upper right-hand edge of the record, but the feebly luminous flame is not visible until it has travelled about 4 cm, after which its speed begins to increase rapidly. From 10 cm the flame speed is approximately uniform until at a point 5 cm from the far end of the tube there is an abrupt arrest. This arrest is distinct from the gradual retardation observed in a long tube, or in a short tube with a much more slowly moving flame, when it is due to the contact of the sides of the elongated flame with the wall of the tube.* That the arrest is due to a different cause in this experiment is indicated by the increased luminosity which spreads backwards from the arrest point through the flame and hot gases behind, evidence of the passage through them of a shock wave. This shock wave is reflected in turn at each end of the tube and still further increases the luminosity of the hot gases, but loses speed after its reflection at the ignition end of the tube.

---

* Payman, *loc. cit.*, fig. 4, Plate 4.

The cause of the arrest is at once apparent from the wave-speed camera record of a repeat explosion of the same mixture, and is seen to be, as suggested by Dixon, a wave reflected from the end of the tube distant from the spark. This record is reproduced in fig. 4, and is explained in the accompanying diagram, fig. 5. It will be seen that the flame, shown in this type of record as a dark line, begins to travel, though extremely slowly at first, immediately the spark has passed. Thus the non-appearance of a record of flame close to the spark in a direct photograph is not an indication of an " induction period," as is sometimes supposed, but merely shows that the slowly moving flame is not actinic enough to affect the photographic film used.

According to the usually accepted explanation, the wave coming from the far end of the tube which caused the arrest of the flame was a sound wave arising originally from the spark. Such a wave would have travelled from the spark to the end of the tube and back to the advancing flame, a total distance of 35 cm, in almost exactly 1 millisecond, and would therefore have had a speed of 350 metres per second, which is appreciably higher than that of sound (about 320 metres per second in this mixture at N.T.P.). The wave from a feeble spark such as was used in these experiments does in fact produce a comparatively gentle shock wave, which moves slightly faster than sound, and this is occasionally registered by the wave-speed camera. Its approximate position is shown by the dotted line in fig. 5, and it is evidently not this wave which causes the retardation of the flame.

The wave sent out by the spark is followed by a continuous series of waves from the flame as it moves forward, owing to the expansion of the burning gases. These waves are feeble as the flame moves slowly but increase in intensity as the flame increases in speed. It is apparently at the point where the flame is accelerating most rapidly that the strong shock wave which later arrests the flame has its origin. The single wave, appearing as a white line in fig. 4, is shown in other photographs to result from the coalescence of a large number of successive waves from the flame front in this region, each apparently travelling at a speed slightly greater than its predecessor.

The apparent speed of this wave decreases as the wave travels towards the end of the tube. It is important here to distinguish between apparent and true speed of the wave. The speed measured directly from the records is of the motion of the wave relative to the wall of the tube and is an " apparent " speed. The " true " speed is the speed relative to the gas mixture. Thus, if a sound wave were moving at a " true " speed of 330 metres per second in air flowing in the same direction along a tube at a speed of 100 metres per second, its " apparent " speed would be 430 metres per second. The speeds of waves given in this paper are " apparent " speeds.

Loss of energy by the wave is insufficient to account for any reduction in its true speed over the short distance measured, and indeed later waves should reinforce it and make it move faster. The change in speed observed is no doubt due to change in motion of the gas through which the wave is moving. The sound wave shown as a dotted line in fig. 5 marks at first the boundary of disturbed gas; behind it as far as the flame front the gas is moving to the left and the speed of its movement will be added to that of the wave moving in it. When the shock wave overtakes the sound wave, the gas ahead of it is no longer in motion and the wave now travels at a lower apparent speed equivalent to its true speed in the gas mixture. The wave travels at an approximately uniform speed of 550 metres per second from a point not far distant from that at which the shock wave overtakes the sound wave.

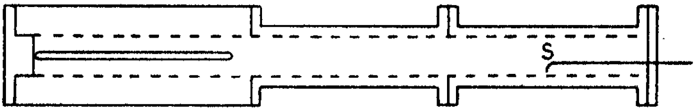
On reaching the end of the tube the shock wave is reflected; it is still a compression wave but as it is moving in the opposite direction it is recorded as a black instead of a white line. It will be seen that the assumption made by earlier workers that the speeds of incident and reflected waves are the same is incorrect. It is probable that the shock wave loses energy on collision with an obstruction, which would account for the loss in speed either partly or wholly. Further, the reflected wave is moving backwards in gas which is being moved forward after the main wave has passed through it. The latter effect can only be small since the gas mixture must be stationary at the end of the tube and there is little sign of diminution in speed as the wave moves away from the end into gas capable of motion. That these effects are operative is shown by the fact that the reflected wave moves backwards at an apparent speed of 310 metres per second, which is less than the speed of sound in the gas mixture.

On reaching the flame and passing through it into the hot gases behind, the speed of the wave increases abruptly from 310 to 1380 metres per second. The marked increase in luminosity then observed in the hot gases, as shown in fig 3, has been generally accepted as proof that these gases are " still burning," the intensity of combustion being supposed to be increased as the pressure is raised by the shock wave passing through them. There are several reasons for doubting this conclusion. It is difficult to imagine that combustion, incomplete in the flame-front, can persist during the length of time indicated by these records. The luminosity does not seem to vary as the wave moves along, whether it is travelling through gas only just traversed by flame, or through gas around the spark which had begun to burn much earlier; further, the increase in luminosity occurs not only in the wave but in the gases behind the wave. If the intensity of the illumination be taken as a criterion of the amount of combustion, there should be more combustion taking place behind the wave than there had been in the flame; and there should be more still after the wave has been reflected from the ignition end of the tube and as successive reflected waves pass backwards and forwards through the mixture. It would appear probable, therefore, that the re-illumination is a result of compression by the shock wave and is not necessarily accompanied by appreciable further combustion or generation of energy. The gases are already hot before the wave passes through them; a very little further combustion therein might result in a great increase in luminosity.

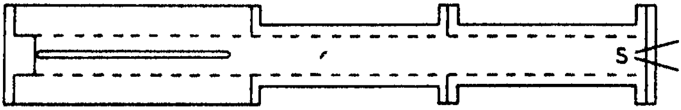
Up to the time when the first shock waves appear in the photographs on Plate 8 the total pressure produced must be small, and the fact that their appearance corresponds with the period of acceleration of the flame probably means that the accelerating flame produces them, and that they do not arise from the gases behind the flame-front. In other records, however, stronger waves appear, the origins of which are undoubtedly in the gases behind the flame, the most important being the " initiating wave " which arises from a region closely behind the flame-front. An example is shown later, in fig. 31, Plate 13. The possibility of these stronger waves being due to the sudden release of energy latent in the burnt gases must be considered. Even in the experiment recorded on Plate 8, the increase in speed of the flame and the sudden appearance of a shock wave may be due to such a cause.

The mean speed of the flame from the spark to the point of arrest is 250 metres per second. After its arrest, the flame travels slowly towards the end of the tube. It is arrested again and is finally pushed forward by the later wave from the ignition end of the tube. The behaviour of the waves and flame after the arrest is not very clear in figs. 3 and 4, though it is evident that there are many waves sent out through or from the flame which, after reflection at the end of the tube, retrace their path and retard the flame. The behaviour of the waves and flame at the end of a tube in somewhat similar circumstances is shown clearly in the series of photographs on Plate 9.

The Effect of Waves Meeting, Overtaking, or Passing Through the Flame—Figs. 6 to 11, Plate 9. In order to examine the mutual effect of waves and flame more thoroughly, longer explosion tubes were used, made up by attaching one or more brass extension tubes to one side of the steel section, so that the latter was always at one end. For the wave-speed records reproduced in figs. 6 and 7, the tube was 95 cm long and consisted, as will be seen from the line diagrams, of the steel section with two brass extension tubes. The gas mixture used had the composition $C_2H_4 + O_2$, and is one which does not detonate readily. In the first experiment, fig. 6, ignition was at a point 15 cm from the right-hand end of the tube, whilst in the second, fig. 7, ignition was at the end. The most striking feature of fig. 6 is the low rate of movement of the flame (shown as a wavy black line) near the far end. Though the shock waves in advance of the flame are not very clear in the reproduction, it is evident in the original record, and can be judged from the reproductions of later records in this paper, that they are largely responsible for the initial driving back of the flame. The considerable backward movement of the flame later appears to be due to suction from behind the flame-front, owing to cooling and contraction of the gases first ignited.



VI.



VII.

The transference of the point of ignition to the end of the tube, fig. 7, produces a flame which moves rapidly over the last section of the tube. The waves ahead of the flame are well marked in this record. The stronger waves which overtake and pass through the flame increase its speed, although at several points the flame is retarded and turned back after encountering reflected waves. On entering the field of view at the upper right-hand side, the flame has a speed of 370 metres per second, the wave emerging from it at the 10-cm mark and shown as a white line having a speed of 450 metres per second. The first wave at the top of the record, only visible during its later stages, is travelling at approximately 450 metres per second; it is a compression wave, being shown as a white line, and is reflected as a compression wave at the end of the tube, changing from white to black when it changes its direction. It will be seen that the reflected wave strikes the flame and retards it at a point just before the shock wave from behind the flame passes through the surface. The reflected wave, from analogy with that in fig. 4, will continue to travel through the flame at an enhanced speed, but this photograph indicates that the wave is also reflected at the flame surface. As the flame surface is the boundary of gas at higher temperature and, presumably, lower density, the reflected wave should be a rarefaction wave, and in agreement with this it is noted that the trace of the wave remains black. A rapid succession of similar traces, apparently also due to rarefaction waves, is to be seen a little later on the record at the point where the flame remains almost stationary for a time. The existence of shock waves of rarefaction is, however, contrary to the laws of thermodynamics, though the possibility that analogous sound waves of very short wave-length may exist for a brief period, rapidly attenuating and becoming true sound waves, has been suggested to us by Professor Jouguet.* In accordance with this suggestion, it is noted in this record that the trace of each of the later waves dies away as it progresses, owing presumably to the density change becoming insufficient to cause refraction of the Schlieren illumination.

Other explanations may also be considered. The fact that the trace of the wave does not change character on reflection could be explained by a change in shape of the wave front, a convex-fronted wave being reflected as a concave-fronted wave. Such a change of shape could be due to the reflecting surface of a stationary flame or of one moving backwards being different from that of a forward-moving flame. The work of Wood† has shown, however, that a concave-fronted wave would only persist for an extremely short interval, after which it would revert to its normal plane or slightly convex-fronted shape.

The first waves to appear show only a slight falling off in speed, indicating that at this stage there is relatively little movement in the gases ahead of the flame. The traces of later waves, however, appear as curved lines, and indicate movement of the gases in which they are travelling.

---

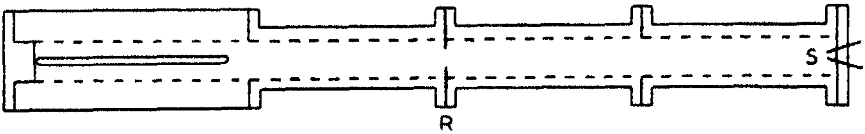
* Private communication.

† 'Phil. Mag.,' vol. 48, p. 218 (1899), and vol. 50, p. 148 (1900).

The waves reflected from or emanating from the flame-front whilst it is almost stationary, immediately after the first arrest, increase in speed as they pass towards the end of the tube. The apparent change in speed of these waves is much more evident after the second arrest, when the flame-front is being pushed back fairly rapidly. The increase in speed of the waves as they travel towards the end of the tube must mean that the gas immediately in front of the flame is moving backwards with it, and that this backward motion decreases the farther the distance from the flame-front. On the other hand, when the flame is again moving forward rapidly after the first arrest, the compression wave reflected from the end of the tube is seen to be retarded a short distance in front of the flame, showing that the gas there is now moving towards the end of the tube. This effect is also shown in later photographs.

The forward movement of the flame after the first arrest appears to be aided by a shock wave or series of shock waves which overtake and pass through it. Reaching the end of the tube first, these waves are reflected and meet the advancing flame, once more retarding it, until it is again pushed forward to the end of the tube by further shock waves passing through it.

This oscillatory motion of the flame appears to have points of similarity with that of the vibrating flames of methane and air or acetylene and air photographed by Mason and Wheeler.* A regularity of vibration similar to that in their photographs was not reproduced near the end of the tube with ethylene-oxygen mixtures, but more regular vibrations were obtained with mixtures of carbon monoxide and oxygen.



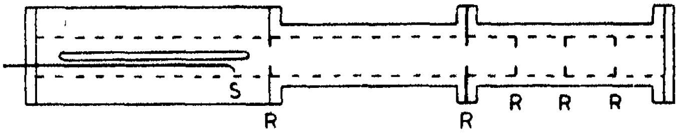
VIII, IX.

Vibrations in the Flame-Front—The flame in a $2CO + O_2$ mixture was found to be too slow in the tubes available to give satisfactory records for our purpose. A faster flame was obtained by using a tube 123 cm long and inserting a small restriction half-way along it. This restriction consisted of an annular brass ring 2 mm thick which reduced the diameter of the tube at that point to 1·9 cm. This method of accelerating the flame will be discussed in more detail later. A direct photograph of the flame travelling along the last section of the tube is reproduced in fig. 8 and a corresponding wave-speed camera record in fig. 9. The photographs are not identical, but they are sufficiently alike to show that the forward movement in each vibration is due to a wave from the hot gases assisting the flame, passing through and leaving it behind, whilst the backward motion is due to the collision between the flame and this wave after its reflection at the end of the tube in the manner suggested by Dixon.*

---

* 'Trans. Chem. Soc.,' vol. 115, p. 578 (1919), and vol. 117, p. 36 (1920).

Passage of a Wave Through the Flame-Front—In fig. 4, Plate 8, a wave meeting a flame was found to retard it and itself to pass through the hot gases behind at a greatly enhanced speed. In order to examine the reverse effect of a wave passing through the flame from the hot gases behind, an experiment was carried out with several restrictions in the tube, and with ignition at some distance from the end. The tube used was 98 cm long, the steel section again forming one end. Four restrictions, each consisting of an annular brass ring 2 mm thick and with a central hole 1·3 cm in diameter, were placed at intervals of 7·5 cm from each other, starting at the end remote from the steel section, and another similar restriction was placed 60 cm from the end, as shown in the diagram. The gas, which in this experiment was a mixture consisting of 95% $2CO + O_2$ and 5% $2H_2 + O_2$, was ignited at a point 30·5 cm from the far end of the steel section. With this arrangement the far end was not flush with the end of the windows as in earlier experiments.



X, XI.

Under these experimental conditions, the sequence of events is as follows. A flame travels outwards from the spark in each direction, slowly along the shorter limb of the tube (the steel section) and more rapidly along the longer limb towards the series of restrictions, so that it reaches these an appreciable time before it reaches the far end of the windows. On reaching the restrictions, the flame is rapidly accelerated and powerful shock waves are sent back along the tube past the spark gap; finally, these shock waves overtake the slowly moving flame in the steel section and pass through it, as shown in the direct photograph, fig. 10, and the wave-speed photograph, fig. 11. The speed of film for these photographs was five times as fast as for figs. 8 and 9. The first slow flame is travelling at a fairly constant speed of about 11 metres per second. The shock wave which overtakes it moves in the gases behind the flame-front at a speed of 1140 metres per second, but its speed changes abruptly to 570 metres per second when it passes into the cold unburnt gas. At the same time a wave, reflected from the flame surface, passes back through the hot gases at a speed of 740 metres per second. As this reflection takes place at a surface where the density of the medium is increased, the reflected wave is a compression wave. The main wave, after passing through the flame, is reflected at the end of the tube (in these photographs 3 cm beyond the field of view), and travels backwards at a speed of 450 metres per second until it reaches the flame, when it travels on through the hot gases at the enhanced speed of 990 metres per second. A second wave arising from the combustion in the restricted zone travels through the hot gases behind the flame at a speed of 1070 metres per second and then, after passing through the flame, travels at 400 metres per second through cold gas beyond. By a coincidence, the backward reflected wave from the end of the tube meets the flame at the instant the second forward wave from the restrictions strikes through. The part of the latter reflected at the flame-front is added to the wave passing into the hot gases, and this wave travels backwards at 990 metres per second, appreciably faster than the first single wave travelling in this direction through the hot gases.

---

* *Loc. cit.*, p. 344.

When the first wave travelling from hot to cold gases passes through the flame-front it accelerates the flame, but the speed attained is much lower than that at which the wave advances. Any acceleration due to the second wave passing through the flame is countered by the effect of the reflected portion of the first wave, so that the flame-front remains almost stationary until driven along again by a later wave. The increased luminosity of the flame or hot gases each time a new wave passes through them is well defined.

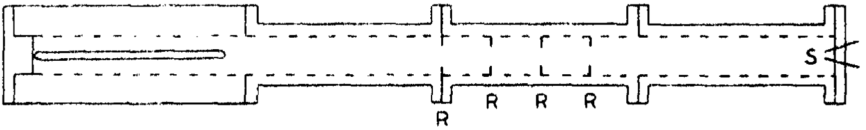
This pair of photographs illustrates the limitations of direct photography. Thus from fig. 10 it might be incorrectly concluded that the first wave that passes through the hot gases is totally reflected at the flame-front, whereas fig. 11 shows that it is only partially reflected, the main portion of the wave proceeding through the flame-front into unburnt gas.

Auto-ignition Ahead of the Flame —Figs. 12 to 17, Plate 10, figs. 18 to 21, Plate 11. That auto-ignition could take place ahead of the flame was first shown by Bradshaw,* and has been confirmed by many other workers. It is most readily obtained at the end of the explosion tube. To investigate its production in this manner an explosion tube 125 cm long was used and restrictions were placed at points along the tube found by experiment to give the desired effect, the restrictions consisting of brass rings 2 mm thick with a central hole 1·3 cm in diameter.

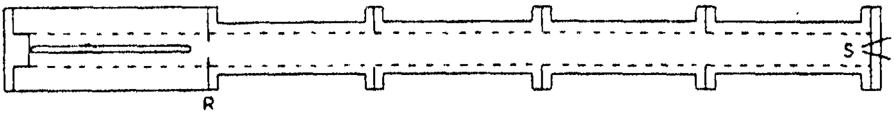
---

* ‘Proc. Roy. Soc.,’ A, vol. 79, p. 236 (1907).

For figs. 12 and 13 four restrictions were placed 7·5 cm apart, the first being at a point 37·5 cm from the spark gap which was at the end of the tube remote from the steel section. The mixture used was $2CO + O_2$. Two shock waves produced as the flame passed through the restrictions can be seen at the top of fig. 13. The two waves this time do not coalesce, presumably because they have either different intrinsic speeds or different radii of curvature. When the first wave reaches the end of the tube, the increase of pressure on impact increases the temperature sufficiently to cause ignition of the mixture and a new flame moves comparatively slowly backwards. The reflected wave travels backwards in front of this second flame until it meets the original flame advancing towards the end of the tube and pushes it back. The unburnt gas between the two flame-fronts is traversed by a large number of waves moving in either direction. The two flames ultimately coalesce.



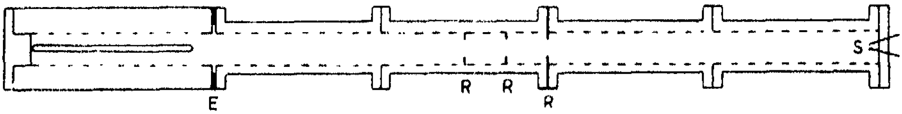
XII, XIII.



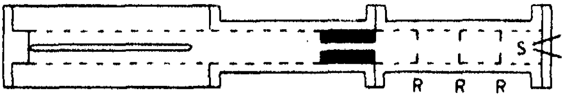
XIV, XV.
XVIII, XIX.

For the experiments recorded in figs. 14 and 15 an addition of 5% of electrolytic gas was made to the mixture $2CO + O_2$, the length of the tube was increased to 155 cm, and one restriction was placed at a point 117 cm from the spark gap. There is a similar sequence of events, but a new feature presents itself, in the region between the two flame-fronts. The reflected wave, as it travels back from the first point of auto-ignition at the end of the tube, ignites the gas mixture at several other points. The mixture in this region has been raised in temperature not only by waves which have passed through it but also by adiabatic compression owing to its own motion towards the end of the tube, and a number of tongues of flame can be seen to travel backwards in the region between the two main fronts.

Similiar auto-ignition ahead of the flame may be produced in gas in "pockets" in the wall of the tube, as shown in figs. 16 and 17. For these experiments the tube was 155 cm long with restrictions placed 61, 68·5, and 76 cm from the spark gap, and for a length of 2 mm, between the end of the steel section and the end of the adjacent extension tube, the diameter of the tube was increased by 6 mm by inserting a brass plate 2 mm thick and with a central hole 3·15 cm in diameter. The mixture $2CO + O_2$ was used. When the first wave reaches the enlargement of the tube, part of it passes into the pocket and ignites the gas there, whilst the main portion proceeds down the tube. The flame so produced, swept forward in front of the main flame, is shown clearly in the records, increasing in size as it spreads backwards and forwards into unburnt mixture; in this manner the "tongue" of flame shown in the photograph is produced. There are other ways in which such tongues of flame may be initiated ahead of the main flame-front and not at a closed end. Figs. 18 and 19, Plate 11, show an interesting example. The experiments were carried out in the same manner as those recorded in figs. 14 and 15 but 6% of electrolytic gas was added to the mixture $2CO + O_2$. In these records the main flame-front has a very broken outline, a number of tongues being produced slightly in advance of the flame as it progresses. It will be seen from the Schlieren record, fig. 19, that each of these tongues of flame has its origin in the powerful shock wave moving just in advance and slightly faster than the main flame. This wave later causes ignition at the closed end, and its intensity is shown by the luminosity and the high speed of the flame produced (*cf.* figs. 13 and 15).



XVI, XVII.



XX, XXI.

For such auto-ignitions ahead of the flame a high speed of flame is not necessary, as is shown in figs. 20 and 21, the film speed for which was only one-fifth of that for figs. 18 and 19. The mixture $2CO + O_2$ was used and the experimental arrangement was as follows: the tube was 95 cm long and ignition was at the end distant from the steel section; three restrictions, each reducing the diameter to 6 mm, were arranged along the first section of the tube so as to cause a rapidly moving flame, and then a section of tubing, 6 mm in internal diameter and 10 cm long, was inserted. Along this narrow tube detonation may be presumed to be set up, but this would die down immediately on reaching the wider tube beyond. The tongue seen just before the second arrest in the photograph in fig. 20 is similar to that shown in fig. 16, although in this experiment there was no " pocket " at the point of origin of the tongue. From the Schlieren records it would appear that the gap between the two flames is filled with hot though not burning or luminous gases, and that ignition ahead is largely due to some turbulence or eddy effect throwing a portion of the flame in advance, though here again the tongue has its origin in the front of the shock wave. Identical direct and wave-speed records were obtained in repeat experiments.

The Movement of Gas Ahead of the Flame—Figs. 22 and 23, Plate 11. It is only in rare instances, for example in fig. 7, that the earlier photographs show any indication of appreciable movement in the gas mixture ahead of the flame, as indicated by variations in the speed of the shock waves therein. With faster flames the indications may be marked, as in figs. 22 and 23. For the record shown in fig. 22 the mixture $2CO + O_2$ was ignited by a black powder fuse at the end distant from the windows of the tube, which was 155 cm long and had no restrictions. The first wave reflected from the end of the tube is seen to become almost stationary relative to the wall of the tube just before meeting the advancing flame. Before being retarded, the wave was travelling at approximately the speed of sound in $2CO + O_2$ (330 metres per second). The shock wave must retain this speed relative to the gas through which it is passing, so that the gas mixture immediately ahead of the flame must be moving at a speed of 330 metres per second. The speed of the flame, which is 610 metres per second, is much greater than the speed of sound in the cold gas.

That the shock wave when apparently stationary has a true motion relative to the gases is shown by what happens to it when it passes into the hot gases behind the flame. On account of the high temperature therein, the wave immediately begins to travel at a speed much higher than that at which the gases are moving in the opposite direction; hence it resumes its travel backwards relative to the wall of the tube.

From the curve of the trace of the wave just after it enters the hot gases, it is evident that there is a forward motion of the gas behind the flame-front which is apparently greater than that of the gas in front of the flame. Since the gases on burning in the flame are increasing rapidly in temperature, this state of affairs can only mean that there is a region just in front and just behind the flame in which the pressure is much higher than it is elsewhere. It is noteworthy that the region immediately behind the flame in fig. 22 shows increased luminosity. The bright band parallel with the flame-front may be due, however, to the flame being convex-fronted, and may be a record of one side of such a flame; but the band only appears with the more rapidly moving flames. From the slope of the trace of the wave it is evident that the marked movement of gas begins about 15 cm in front of the flame and continues for at least 4 or 5 cm behind it. At the point where the apparent change in speed of the wave is noted, a number of waves appear and these are evidently responsible for the driving forward of the gas ahead of the flame. It must be noted that a wave moving in one direction always falls in speed after collision with another wave moving in the opposite direction, and the diminution in speed of the reflected wave we have been discussing will be partly due to the effect of collision with forward moving waves. The waves appear to pass through each other, but it is possible that they are reflected on collision. Fig. 22 proves that a constant speed of flame does not of necessity mean quiescence in the gas ahead of the flame, as has been suggested by Bone and Fraser.*

A still more striking example of the effects of movement of the gas on the speed of the wave is shown in fig. 23, where a reflected wave from the end is caused to reverse its direction completely and is driven back so that it again strikes the closed end. The mixture used for this experiment contained 45% of ethylene and 55% of oxygen, and was ignited by a spark at the end of a tube 95 cm long without restrictions. The wave driven back in fig. 23 reaches the end for the second time at approximately the same time as the flame, and as a result an extremely rapid and powerful wave travels back through the hot gases, rendering them intensely luminous. The phenomenon observed here is similar to that observed by Dixon, who noted that an extremely violent " retonation " wave was sent back along a tube if detonation was set up at the end of the vessel immediately before the flame arrived there.

The Initiation of Detonation in a Short Tube—Figs. 24 to 29, Plate 12. We have so far considered only the behaviour of waves appearing in conjunction with flames other than those of detonation or giving rise to detonation. The initiation and incidence of detonation were examined in the first place with strong mixtures in the shortest tube, namely, the steel section only, as for the experiments recorded on Plate 8.

---

* 'Phil. Trans.,' vol. 230, p. 363 (1931).

With the mixture $C_2H_4 + 3O_2$ it has been shown* that detonation is set up after a run of only 6 cm in a tube 1 metre long, and an almost identical result was obtained in the present series of experiments with the tube only 30 cm long. Slight changes in the composition of the mixture caused the starting-point of detonation to recede farther and farther from the point of ignition. Fig. 24 is a wave-speed record with the mixture $C_2H_4 + 4O_2$. In this, as in other wave-speed records on the same Plate, the shock wave from the spark, as shown in broken line in fig. 5, is recorded distinctly, and a number of shock waves are sent out in advance of the flame in addition to the wave sent out when the flame accelerates. When comparing this record with that in fig. 4 it must be remembered that the flame there is travelling more slowly, the record being obtained with a film moving at a lower speed. In fig. 24 the first waves on reaching the end of the tube are evidently intense enough not only to ignite the gas mixture but to initiate detonation instantaneously therein. The detonation wave starting at the end of the tube is very brilliant, and the photograph is evidently affected by halation, but the continuation of the detonation wave as a shock wave in the hot gases appears to take place with little change in speed, though with diminution in luminosity.

With a rather more dilute mixture, containing 18% of ethylene, the shock waves striking the end of the tube in fig. 25 were evidently not powerful enough to ignite the gas there, but on collision with the flame after reflection at the end sufficient further rise of temperature appears to have been caused to initiate detonation in the flame-front.

With a mixture containing 30% of ethylene detonation is set up in a manner which in our experience is the most usual, that is to say, in the flame-front without the aid of reflected waves or the impact of waves with an obstruction such as a closed end. A direct photograph is reproduced in fig. 26 and a wave-speed record in fig. 27. It is evident that the gas mixture in this experiment has been prepared for detonation before the flame reaches it, that is to say, it has been pre-heated temporarily by a number of shock waves which have passed through immediately in advance of the flame, and more permanently by adiabatic compression. It will be noted that the wave reflected from the end of the tube moves more rapidly back through the hot gases than the retonation wave, presumably because the gas mixture through which the reflected wave is travelling has been raised to a higher temperature by the passage of the retonation wave through it. The uniformity in speed of the reflected wave appears to negative the alternative suggestion that it is travelling through gas set in motion by the retonation wave.

---

* Payman *loc. cit.*, fig. 10.

A combination of the effect of waves sent ahead of the flame before and after reflection at the end of the tube is shown in the direct photograph fig. 28, and the wave-speed record fig. 29. An 18%-mixture of ethylene and oxygen was used, but in these experiments sufficient increase in temperature was evidently caused by the collision of normal and reflected waves moving in gas already heated by adiabatic compression to ignite the gas mixture ahead of the flame, and to cause immediate detonation. Though this possibility was doubted by Dixon, it has been proved by the experiments of Campbell and Woodhead,* Egerton and Gates,† and Bone and Fraser.‡ It would not appear necessary to stipulate the presence of reflected waves, for the rise in temperature might equally well be due to concentration of pressure caused by one powerful wave overtaking another; it has not been found possible, however, to reproduce these conditions in the apparatus at our disposal, though the tongues of flame seen in figs. 18 and 19 may have been caused in this manner.

The bright band immediately behind the flame front is again noticeable in the direct photographs, figs. 26 and 28.

The Initiation of Detonation in Longer Tubes—Figs. 30 to 34, Plate 13. The setting up of detonation has been examined further under various conditions in a longer tube. For fig. 30 a mixture containing 50% of ethylene was ignited in an unrestricted tube 95 cm long, the last 30 cm being photographed. Here again detonation is set up in the flame-front, which is passing through gases which have been heated both by shock waves slightly in advance and by adiabatic compression. Reflected waves are also much in evidence; the third may have contributed to the setting up of detonation here, though the two earlier waves have had the effect of slowing down the flame. The retardation of the reflected waves as they approach the flame is very clearly marked, as is also the loss in speed of the incident waves as they travel from the flame to the end of the tube, an effect due partly to collisions of waves passing in opposite directions and partly to the forward movement of the gas, which becomes less and less as it approaches the end of the tube.

---

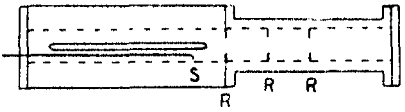
* ' J. Chem. Soc.,' p. 3010 (1926).

† ' Proc. Roy. Soc.,' A, vol. 114, p. 137 (1927).

‡ ' Phil. Trans.,' A, vol. 228, p. 228 (1929).

In order to obtain initiation of detonation further from the end of the tube, a more strongly explosive mixture was used, containing 45% of ethylene, and an extra 30-cm length was added to the explosion tube. Direct and wave-speed records are reproduced in figs. 31 and 32. Though there is evidence of waves passing through unburnt gas mixture in front of the flame before detonation is set up, the detonation in this experiment has its origin behind the flame-front, the " initiating wave "* being clearly seen, and the detonation and retonation waves start forward and backward from the point where the initiating wave overtakes the flame-front. The initiating wave appears to have its origin in that region behind the flame-front where it has been shown there is a marked local increase in pressure.

The effect of a deliberately induced initiating wave is shown in figs. 33 and 34. The experimental method used for figs. 10 and 11 was used, with a more explosive gas mixture containing 40% of ethylene. The tube used was 67 cm long; it consisted of the steel section and one brass section 30 cm long. The brass section contained three restrictions, placed 15 cm, 22·5 cm, and 30 cm from the end, each restriction having a central hole half the diameter of the tube. The gas was ignited at a point 30 cm from the end of the steel section; this end did not correspond with the end of the windows, but was 3 cm beyond them. The spark gap was intentionally placed away from the axis of the tube so as not to interfere with the light transmitted through the windows. The beginning of the flame is therefore not recorded, but shortly afterwards the flame can be seen to be moving outwards towards each end of the tube at approximately the same speed, the equality being due to the fact that the point of ignition is nearly midway between the two ends of the tube. One portion of the flame proceeds smoothly along the steel section, but when the other passes into the restricted portion of the brass tube, a powerful shock wave is set up which travels back and overtakes the flame-front in the steel section, instantly setting up detonation there. A short initiating wave can be seen to develop just behind the flame-front immediately before detonation is set up.



XXXIII. XXXIV.

This experiment demonstrates once more the ability of the " initiating wave " to overtake the flame and set up detonation.

---

* Payman, *loc. cit.*

The Initiation of Detonation by Means of a Detonator—Figs. 35 and 36, Plate 14. The next series was carried out some time ago as a continuation of earlier experiments, in an unsuccessful attempt to find a simple method of measuring the "strength" of a detonator. It was thought possible that the distance taken for detonation to be set up in a gas mixture which detonated with difficulty, such as carbon monoxide and oxygen, might be related to the strength of the detonator. The mean initial speed of the shock wave sent out by a detonator through the atmosphere had been found* to vary but slightly with detonators of different strengths, increasing to a maximum as the weight of fulminate was increased, when further increase caused no alteration in the speed of the wave. In these experiments, the mixture used was of composition $2CO + O_2$, saturated with water vapour at 10° C and was contained in a glass tube 1·3 cm in diameter and 1 metre long. In order to measure the "end effect," each detonator was placed in a central hole in a rubber stopper at one end of the explosion tube, so that only the end of the detonator was in contact with the gas mixture. The first 80-cm length of the tube was photographed by direct photography, using a film speed of 70 metres per second.

It was found, contrary to the statement of Le Chatelier,† that detonation was set up instantaneously, and always at a rate higher than the normal, with any of the usual sizes of commercial copper-cased fulminate-chlorate detonators and with a No. 6 aluminium-cased azide-tetryl detonator. The tendency of the detonation wave to have an abnormally high speed immediately after initiation has been frequently observed.

TABLE I

Detonator No.	Rates of detonation in metres per second			
	0–20 cm	20–40 cm	40–60 cm	60–80 cm
3	2020	2000	1900	1810
4	2000	1900	1870	1820
7	1950	1890	1770	1740
8	1940	1860	1770	1740
6 (azide-tetryl)	2480	2400	2180	1940

The mean speed of the wave over successive 20-cm lengths of the tube with each detonator is given in Table I. It will be seen that the rate is higher in each section the weaker the detonator used. In no instance did the speed of the wave fall below the rate of detonation, that is the steady rate which is attained after the wave has travelled beyond the

---

* Payman and Shepherd, 'Pap. Safety Min. Res. Bd. Lond.,' No. 29 (1926).

† 'C.R. Acad. Sci., Paris,' vol. 130, p. 1757 (1900).

*sphere of influence of any initial impetus imposed when detonation begins, and which is about 1740 metres per second in this mixture**; only with the two strongest detonators did it fall to this speed even in the last section photographed.

It is evident that the rates give no measure of the relative strengths of the detonators. This is emphasized by the result with the No. 6 azide-tetryl detonator, with which the measured rates are throughout greater than with any fulminate-chlorate detonator. The apparent diminution of initiating effect with the heavier fulminate detonators recalls the analogous results obtained in "gap tests," using an ammonium nitrate explosive as receptor, recorded in Part II.†

Two typical records are reproduced in figs. 35 and 36. Fig. 35 shows ignition by the strongest fulminate-chlorate detonator, No. 8, and the usual striated flame. With ignition by the azide-tetryl detonator, fig. 36, the flame is not striated. As little as 0·5% of hydrogen added to the mixture caused the striations to disappear also when the fulminate-chlorate detonator was used.

The Effect of Restrictions—Figs. 37 to 39, Plate 14. Various methods have been used to influence the initiation of detonation by imposing shock waves on the flame. Perhaps the most easily produced shock wave is that from an electric spark, but the wave from a comparatively weak spark is feeble and soon dies away; with a more intense spark unknown electrical effects may be introduced. Detonators give shock waves which are powerful, but these are not simple waves, being usually affected by the presence of solid particles from the detonator case and the detonating composition, and by the gaseous products of decomposition of the latter. In the present paper it has been thought desirable to limit the shock waves used to those produced by the burning gas mixture itself, so avoiding all external complications, using restrictions in the tube for this purpose.

The use of annular restricting rings was first suggested by Mason and Wheeler,‡ and was shown by them to enhance greatly the speed of flame in the slow-burning methane-air mixtures. We have found that the method can be used for producing powerful shock waves in more rapidly burning mixtures and for initiating detonation very readily in mixtures which only set up detonation with difficulty in tubes of smooth bore. Thus, with the mixture $2CO + O_2$, which normally requires a long run, detonation was set up after a travel of only 60 cm in a tube 2·5 cm in diameter and 95 cm long containing five of the thin annular brass restrictions at 7·5, 15, 22·5, 30, and 37·5 cm from the point of ignition.

---

* Campbell, Whitworth, and Woodhead, 'J. Chem. Soc.,' p. 59 (1933).

† Payman, Woodhead, and Titman, 'Proc. Roy. Soc.,' A, vol. 148, p. 607 (1935).

‡ 'J. Chem. Soc.,' vol. 117, p. 36 (1920).

The effect of a restriction in initiating detonation in a mixture of ethylene and oxygen is shown in fig. 37, for which the experimental conditions were the same as for the record shown in fig. 3, Plate 10, except that a restriction was placed at the centre of the third window; its position may be gauged from the additional thin black line where it cuts out the light in the centre of the left-hand section of the record. The flame starts out from the spark at the same speed until it is arrested, much earlier this time, and apparently by a wave reflected from the restriction. Later, the flame passes through the restriction, increasing in speed as it does so, and a short time afterwards a wave leaves the restriction, travelling in both directions and doubtless initiating detonation in the flame-front as soon as it overtakes it. This powerful wave has its origin at the restriction, presumably because there has been a piling up of gas, and therefore of pressure, with a greatly increased temperature at that point.

Another mode of action of a restriction is shown in the wave-speed record, fig. 38, the mixture being $2CO + O_2$ with an addition of 10% of $2H_2 + O_2$. The tube was 155 cm long and had one restriction, this time in the centre of the right-hand window. A powerful wave is seen to ignite the gas mixture on striking the restriction and secondary flames travel comparatively slowly away, forwards to be later overtaken by the main flame and backwards to meet it. Here again a powerful wave leaves the restriction shortly after the flame has passed through, and initiates detonation on passing through the flame-front.

Fig. 39 affords an interesting comparison. The same tube was used but with the more slowly burning mixture $2CO + O_2$, and a second restriction was placed at the centre of the middle window. Here the accelerating effect of the restrictions is not so great, the interesting point being that the shock wave on reaching the second restriction does not cause ignition but pushes the gas forward in the vortex ring formation studied by Shepherd.*

DISCUSSION

When an inflammable gas mixture is ignited in a tube the behaviour of the flame depends to a large extent upon whether the ends of the tube are open or closed. The present research has been limited to an examination of the propagation of flame and the shock waves produced thereby in a closed tube. In order to examine what happens during the propagation of flame, it will be necessary to discuss the conditions which obtain in cold and hot gases and the production, form, and effect of shock waves in rather greater detail than has been done previously.

---

* 'Bull. U.S. Bur. of Mines, Wash.,' No. 354 (1932).

When a flame travels through a gas mixture in a closed vessel the gas must be set in motion in front and behind in all circumstances other than those of detonation, which we shall neglect for the moment. The hot gases produced by combustion in the flame will expand and push forward the cold gas in front and push backwards the hot gas behind the flame, always tending to equalize the pressure in front and behind. It is clear that this tendency will be more effective the slower the flame. With the slowest flames the relative motion of gas forwards and backwards will depend upon the length of the column of gas in front and behind the flame, for, to obtain an equivalent compression, movement must be greater the longer the column. This is well illustrated by the experiments of Ellis,* who ignited gas mixtures asymmetrically in long closed tubes and found that the two flames so produced, moving in opposite directions, travelled faster along the longer than along the shorter limbs, and always tended to reach the two ends of the tubes at the same instant. In the experiments described in the present paper, the photographs are always taken at one end of the tube, usually with the flame moving towards that end; the amount of motion ahead of the flame will be at a minimum in these circumstances. The motion of the gas will be further complicated by the cooling of the hot gases by conduction to the walls, the resulting reduction in pressure causing a suction effect, drawing backwards the gas behind the flame and sometimes even the flame itself.

This does not imply that a constant speed of flame is not possible for a period in such circumstances, though it is unusual. This is illustrated by the experiments of Stevens,† who ignited a mixture at the centre of a soap bubble; the bubble increased in size as the flame expanded and pushed the gas ahead of it forward, but a constant speed of flame was obtained. A simple calculation enabled Stevens to allow for the motion of the gases ahead of the flame and thus to find the fundamental speed of the flame, that is to say, the speed due to transfer of heat only. It would obviously be difficult to obtain the fundamental speed of flame from any measurements of constant speed in a closed vessel, for too many factors would have to be taken into consideration even if they were all known or measurable. The fundamental speed can, however, be calculated and the necessary allowances made either (1) when the gas mixture is burning in a bunsen burner* or (2) when the flame is travelling from an open to a closed end so that the uniform movement of flame is measured.† In the former method allowance is made mainly for the speed of the gas current; and in the latter allowance is made only for the shape of the flame, there being no measurable forward motion ahead of it, and the motion of the hot gases backwards having no apparent effect, provided that the flame is sufficiently close to the open end to enable them to escape readily into the air. The fundamental speed is lower than the measured speed when the flame is in motion in a stationary gas mixture.

---

* 'J. Chem. Soc.,' vol. 123, p. 1440 (1923).

† 'J. Amer. Chem. Soc.,' vol. 48, p. 1896 (1926).

With slowly moving flames the speed is therefore dependent almost entirely on three factors: (1) the fundamental speed, (2) the shape of the flame, and (3) the speed of the gas mixture. This is no longer true with more rapidly moving flames because (1) conditions may be produced where the pressure is no longer even approximately uniform in the gas mixture, and (2) shock waves may be produced and have a specific effect on the flame. The dividing line was suggested in an earlier paper‡ to be roughly that defined by the speed of sound in the cold gas. Marked variations in the pressure of the gas mixture other than that in the shock waves themselves are illustrated, for example, in fig. 22, Plate 11, and have already been discussed briefly. In fig. 22 there is an increase in pressure immediately in front and behind the flame. This is due to the gas set in motion by the flame being unable to move forward as rapidly as the flame is attempting to push it in virtue of the expansion of the gaseous products of reaction. The piling up of pressure is more readily comprehended if the flame and expanding gases are regarded as a solid piston moving along a closed cylinder. With a slow motion, the gas ahead of the piston is at approximately uniform pressure at any one instant, but the velocity of layers of gas varies gradually from a maximum immediately in front of the piston to zero at the far closed end of the cylinder. With a rapidly moving piston, the gas in the region immediately in front of it will be at high pressure and moving at high speed whilst that a short distance away will be approximately at its original pressure and will be stationary.§

This picture of what happens with a flame moving rapidly along a tube is important, for it implies that with such a flame the gas ahead will have been compressed and so heated above its original temperature. Reaction

* Gouy, 'Ann. Chim. Phys.,' vol. 18, p. 5 (1879).

† Coward and Hartwell, 'J. Chem. Soc.,' p. 2676 (1932).

‡ Payman, *loc. cit.*

§ *See* 'Trans. Inst. Min. Eng.,' vol. 88, p. 459 (1935).

in this mixture may be presumed, in virtue of the higher initial temperature, to be initiated more rapidly; thus we have the necessary additional conditions for the sudden setting up of detonation in advance of or in the flame-front, *for detonation is presumably set up owing to a suddenly increased rate of reaction.* Further, since the compression in the unburnt mixture is confined to layers of gas immediately in front of the flame, the heat of compression is not lost to the wall of the tube before the flame reaches them, as it would be with the slower type of compression. The conditions as the speed increases are also such as to cause a cumulative increase in the flame speed.

Comparatively little is known about the effect of pressure on the speed of flame. The high pressure immediately behind the flame-front will, however, result in a more rapid recombination of ions and a more rapid release of the latent energy remaining in the gases behind the flame. This would possibly account for the appearance of the initiating wave a short distance behind the flame when this energy is released, with the result that detonation is set up as soon as this wave passes through the flame-front; a similar series of events may be responsible for the milder waves which emerge from the flame soon after its inception.

It will be convenient at this stage to consider the formation of these waves and their effect. The form of the wave and the pressure therein are being examined in a separate section of this research, but much information as to their character is to be obtained from the photographic records in this paper. A review of earlier work dealing with shock waves both mathematically and experimentally has been given in an earlier paper.* These differ from sound waves in that they travel faster, are of greater amplitude, and are transient. They mark the boundary of the effect of the disturbance producing them but may travel much faster than it. The rise of pressure in front of the shock wave is almost instantaneous and the heating therein adiabatic, but the pressure is only maintained for an extremely short time when the wave advances ahead of the disturbance, and reduction of pressure and cooling may be equally rapid. Nevertheless, the effect of such waves on flame propagation is important because (1) they will set in motion the gas through which they travel, and (2) they may heat the gas for a sufficient length of time to begin chemical reaction which will be itself more or less self-supporting. The first effect is noticeable in many of the photographs in this paper, notably those on Plate 9; the waves may be seen to accelerate flames which they overtake and retard flames with which they collide. The second effect is most marked when a wave sets up detonation on passing through the flame.

---

* Payman and Robinson, 'Pap. Safety Min. Res. Bd. Lond.,' No. 18 (1926).

An acceleration of the flame forwards and a retardation or reversal may not always be due to a shock wave. In earlier stages the production of pressure by combustion has presumably been greater than the loss due to cooling of burnt gases; when the flame slows down the balance is upset and the cooling gases exert the greater effect. A second mode of arrest, shown more particularly in shorter tubes, and illustrated in the present paper in figs. 3 and 4, is due to a shock wave sent out by the flame; this type of retardation has been frequently observed by other workers. The acceleration of the flame by waves passing through it has also been observed by other workers, for the passage of waves in the hot gases is well marked in direct photographic records. The acceleration of the flame immediately following ignition must also be considered; this is no doubt due in some degree to the increasing size of the flame in its early stages.

The speed of the wave will depend not only on the motion but also, and to a more marked extent, on the temperature and density of the gas through which it passes. The effect of the motion of the gas has been generally overlooked, but that it may be appreciable is shown by figs. 22 and 23, Plate 11. The effect of the temperature and pressure of the gas has also frequently been ignored by investigators, who have assumed, in their explanatory diagrams, that shock waves travel at the same speed in a gas whether it be hot or cold. It is unsafe to assume that waves travelling through hot gases are more intense than those travelling through cold gases because of their high speed and the luminosity they produce. They may be travelling at speeds little greater than that of sound in the gas mixture at the temperature prevailing, implying that their intensity, in comparison with the pressure of the surrounding gas, is not very great. In the cold gas the shock waves travel at speeds much higher than that of sound, so that their intensity may be considerably higher than that of shock waves travelling through hot gases at much higher speeds. Ignition may thus be brought about in shock waves travelling ahead of the flame, especially when they collide or overtake each other, when their amplitudes may be added, or when a wave is reflected from a closed end, when the pressure may be assumed to be momentarily doubled.

Detonation is a limiting condition when the effect of the shock wave is sufficient to cause immediate reaction in the gas through which it passes, the resulting combustion maintaining the original pressure in the shock wave. This implies that the true rate of detonation will be uniform under constant conditions of experiment, and that it will be the maximum stable rate. Nevertheless, the detonation wave on inception usually starts off at an abnormally high speed. There are two possible causes for this. As we have seen, the gas mixture in front of a rapidly moving flame is itself moving forward rapidly and its speed will be added to the speed of detonation immediately after initiation. There will be no movement in the gas ahead of the detonation wave other than that set up before detonation was initiated. Secondly, and this will probably be more important, the energy of waves overtaking the flame will be added to that of detonation, causing an increase in speed until the added energy is dissipated. The latter effect is shown particularly well in the results obtained with detonators as means of ignition. It may (1) assist and increase the rate of a detonating mixture, or (2) cause apparent detonation in a mixture which without it would certainly not maintain detonation.

The origin of the shock waves which have such a marked effect on the speed of the flame is evidently (in the absence of artificially imposed waves) always in the flame-front or a short distance behind it. Though the cause of their production is still a matter of conjecture, their subsequent behaviour (and that of artificially imposed waves) and their effect on the flame are in accordance with the generally accepted views on the detonation of gaseous mixtures.

SUMMARY

A study has been made of the inflammation of certain gas mixtures by means of the wave-speed camera, in order to examine more thoroughly than is possible by direct photographic means the initiation of detonation and the production of shock waves and their effect on the flame prior to the setting up of detonation. The mixtures used were of ethylene and oxygen, which detonate with ease, and of carbon monoxide and oxygen which do so with comparative difficulty.

The speed of flame in a tube, before shock waves of appreciable intensity are produced, depends upon (*a*) the fundamental speed of propagation of flame by the transfer of heat, and (*b*) the motion of the gas in which the flame is moving. Arrest of the flame early on is then due to its sudden cooling when it first touches the wall of the tube; a similar effect may be produced when the flame is retarded near the end of the tube just before its final extinction. When shock waves are formed, they may push the flame forward at a lower speed than that at which they are passing through it, or they may retard the flame on meeting it after reflection from a closed end. Hence in a closed system the flame may be made to oscillate.

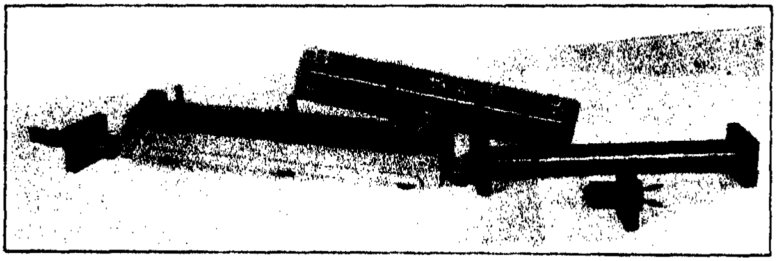


FIG. 2 The explosion vessel

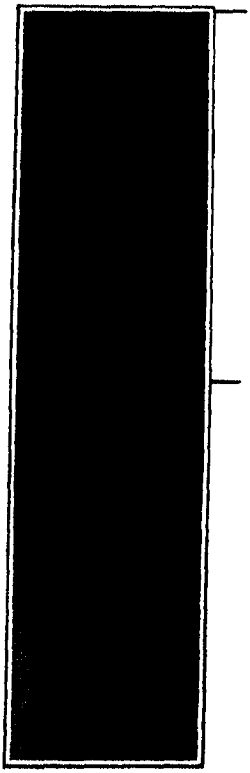


FIG. 3
Direct photograph

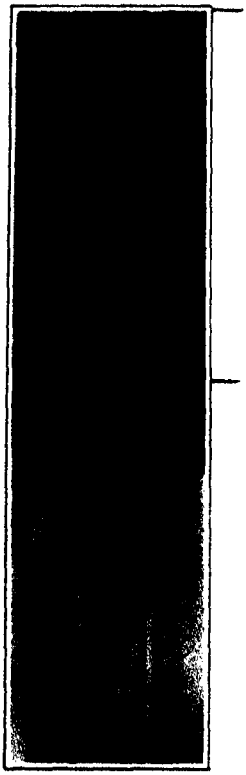


FIG. 4
Wave-speed record

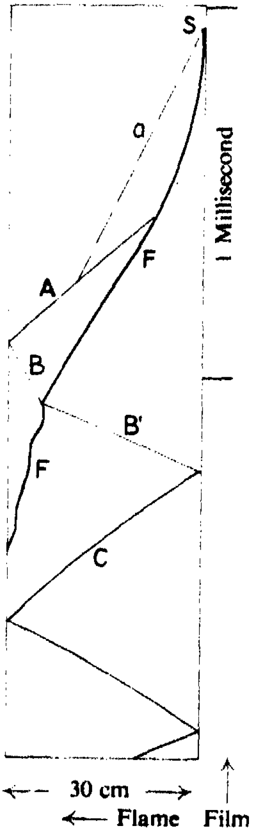


FIG. 5

S	Spark
a	Sound wave
A	Shock wave
B	Reflected wave (from A)
B′	Wave in hot gases
C	Reflected wave (from B′)
F	Flame

Flame and shock waves in a comparatively slowly burning mixture of ethylene and oxygen (15% C_2H_4)

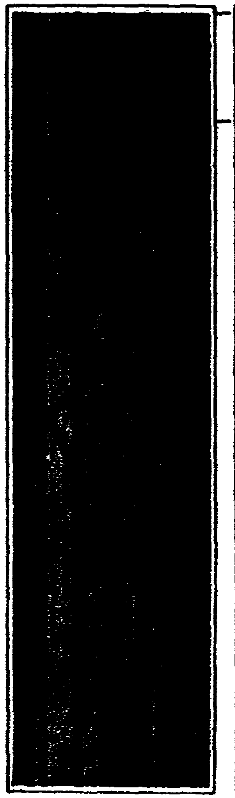


FIG. 6
Wave-speed

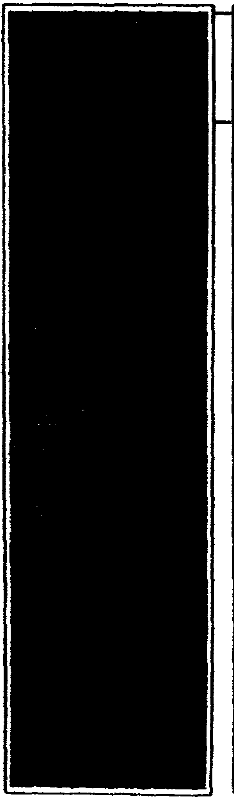


FIG. 7
Wave-speed

$C_2H_4 + O_2$

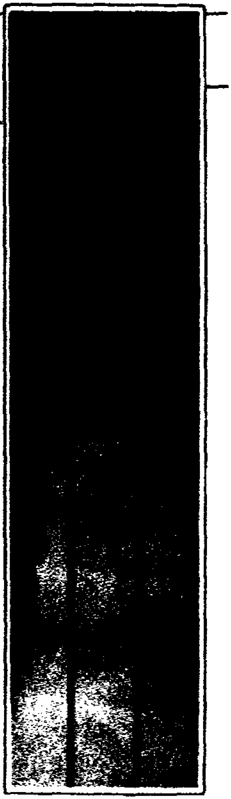


FIG. 8
Direct

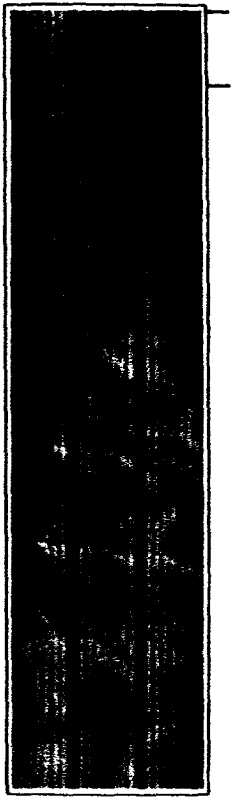


FIG. 9
Wave-speed

$2CO + O_2$

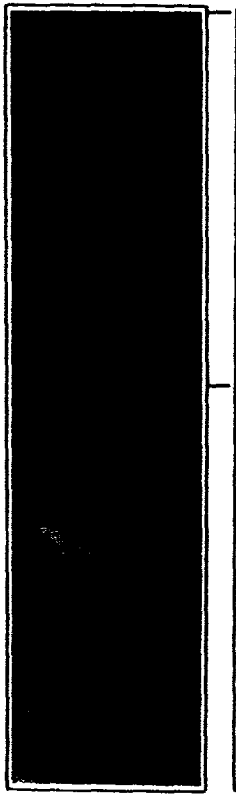


FIG. 10
Direct

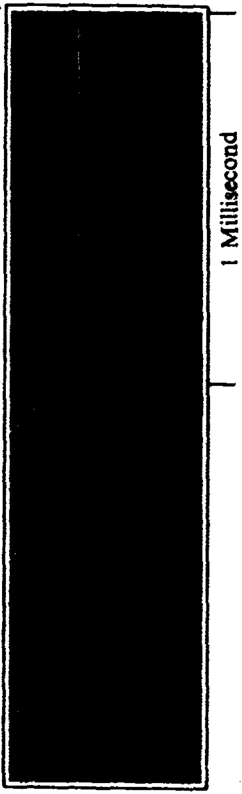


FIG. 11
Wave-speed

$2CO + O_2 (+ 5\% \, 2H_2 + O_2)$

The effect of waves meeting, overtaking, or passing through the flame

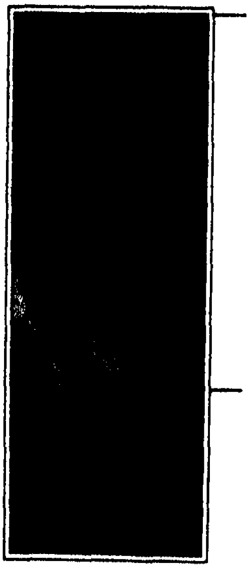


FIG. 12

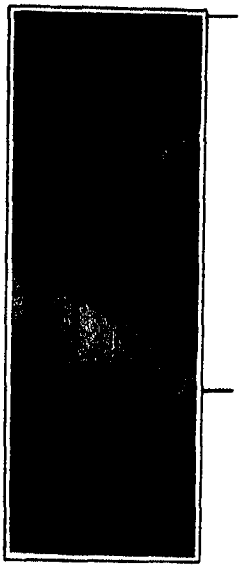


FIG. 14

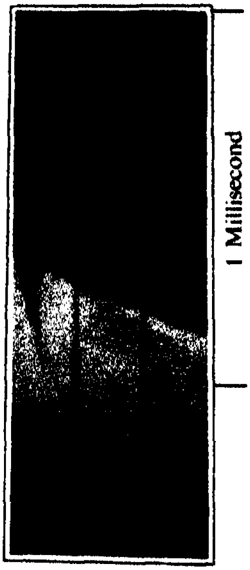


FIG. 16

Direct photographs $2CO + O_2$

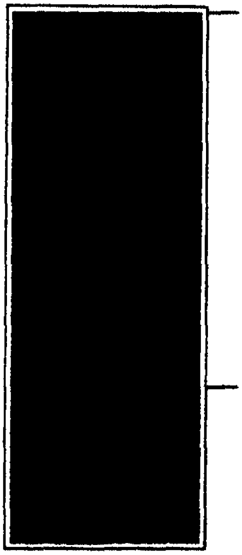


FIG. 13

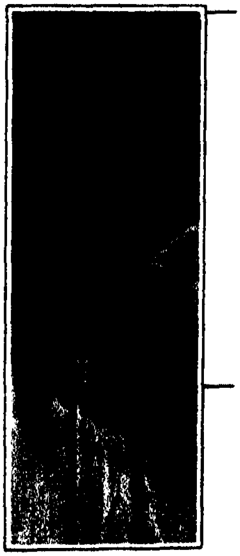


FIG. 15

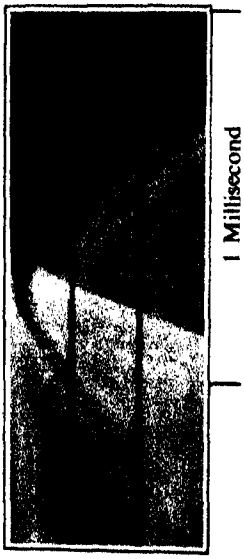


FIG. 17

Wave-speed records $2CO + O_2$

Auto-ignition at a closed end

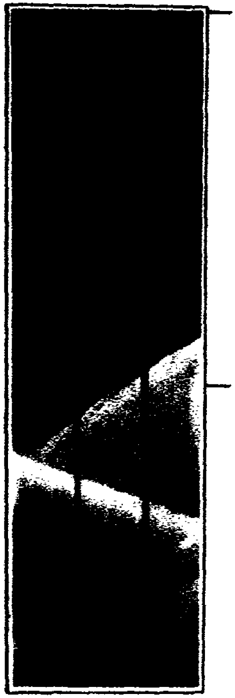


FIG. 18
Direct

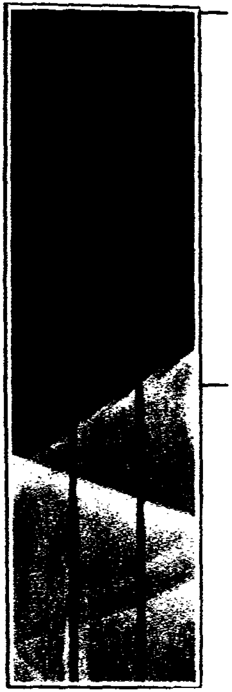


FIG. 19
Wave-speed

$2CO + O_2 (+ 5\% \; 2H_2 + O_2)$

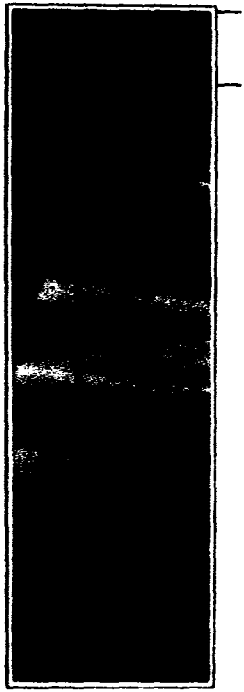


FIG. 20
Direct

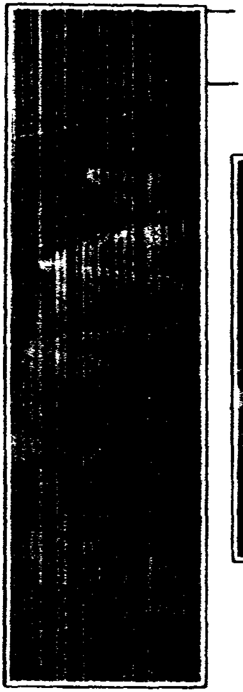


FIG. 21
Wave-speed

$2CO + O_2$

Auto-ignition ahead of the flame

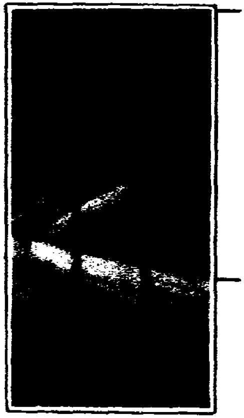


FIG. 22
Wave-speed

$2CO + O_2$

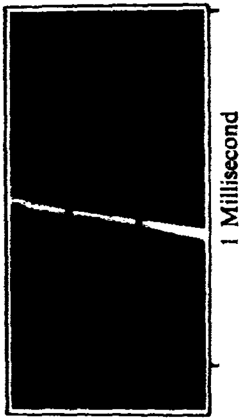


FIG. 23
Wave-speed

$45\% \; C_2H_4$

Movement of gas ahead of the flame

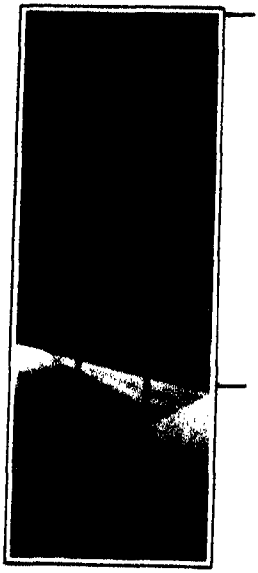


Fig. 24

20% C_2H_4

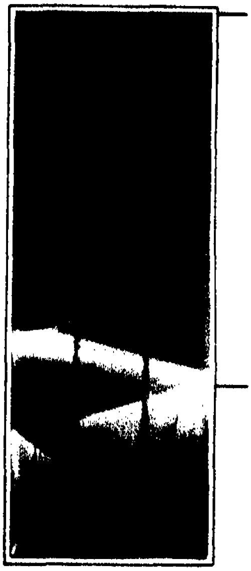


Fig. 26

Direct photographs

30% C_2H_4

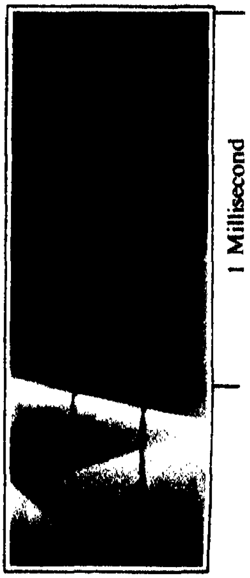


Fig. 28

18% C_2H_4

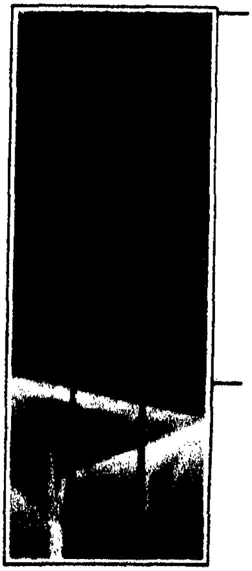


Fig. 25

18% C_2H_4

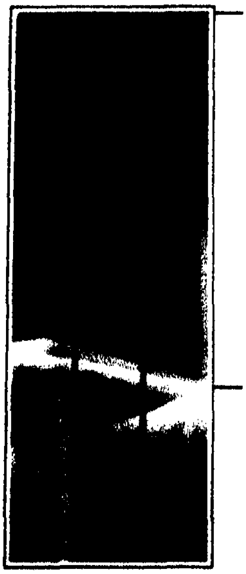


Fig. 27

Wave-speed records

30% C_2H_4

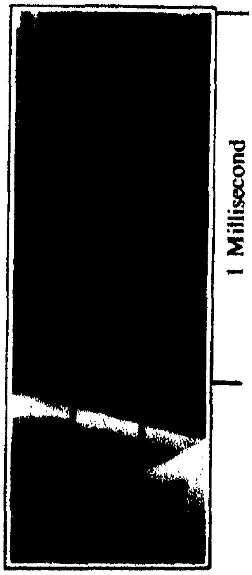


Fig. 29

18% C_2H_4

Initiation of detonation in a short tube

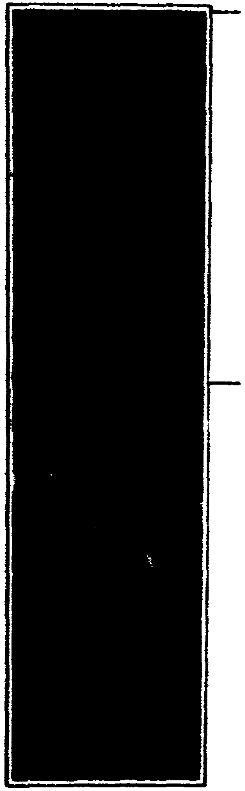


Fig. 30
Wave-speed
$C_2H_4 + O_2$

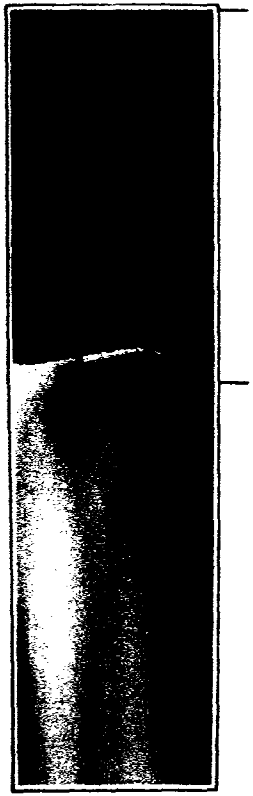


Fig. 31
Direct

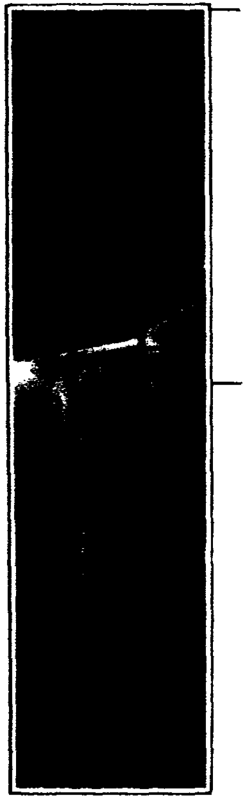


Fig. 32
Wave-speed

$2C_2H_4 + 3O_2$

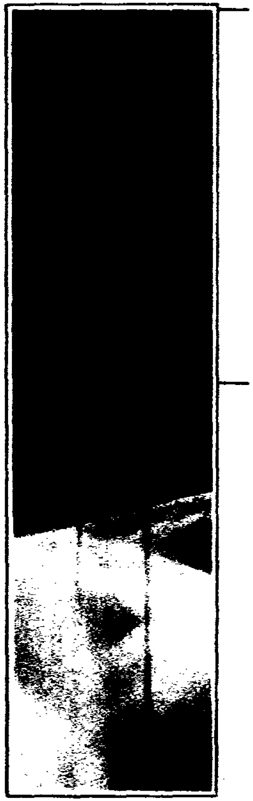


Fig. 33
Direct

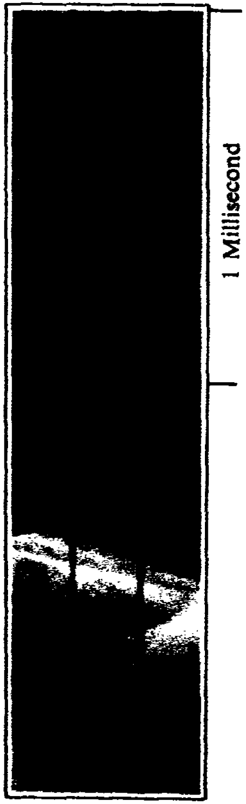


Fig. 34
Wave-speed

$2C_2H_4 + 3O_2$

The initiation of detonation in a longer tube

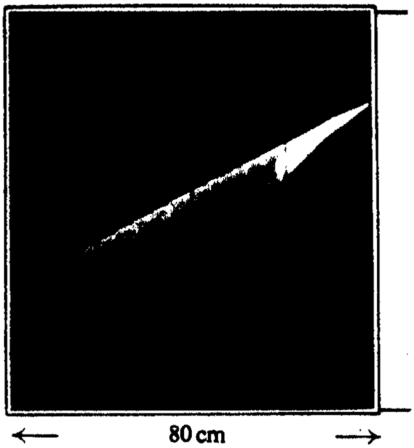


FIG. 35

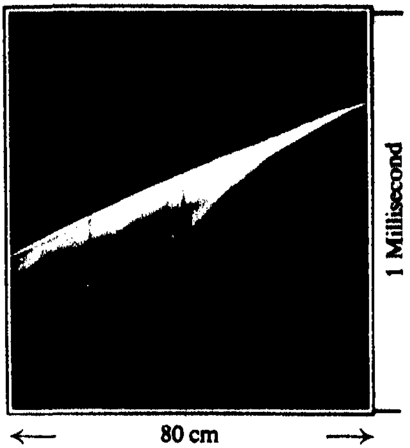


FIG. 36

Direct photographs $2CO + O_2$

Initiation of detonation by means of blasting detonators

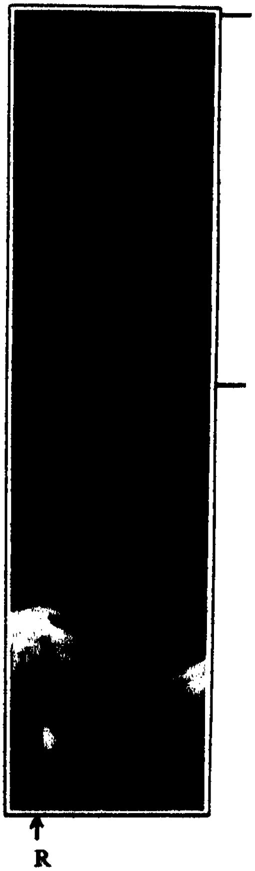


FIG. 37

15% C_2H_4

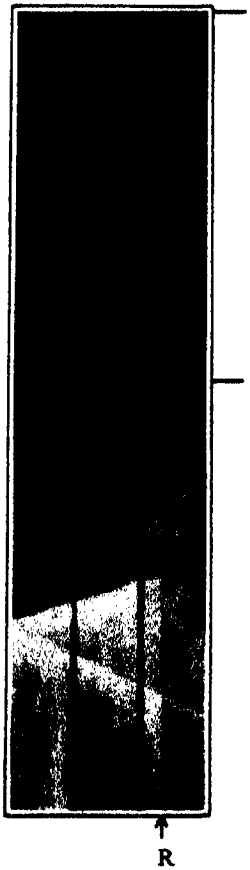


FIG. 38

$2CO + O_2$ (+ 10% $2H_2 + O_2$)

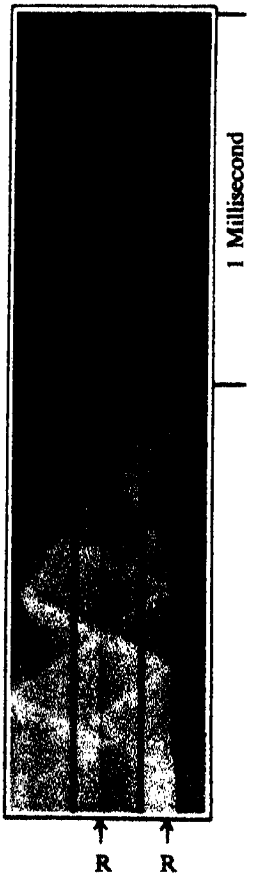


FIG. 39

$2CO + O_2$

The effect of restrictions

The effect of the wave from a spark of moderate intensity is negligible; the wave formerly assumed to be from the spark, and known to have a pronounced effect on the flame, arises in the flame, or possibly in the region immediately in the rear of the flame, some time after ignition.

Shock waves do not necessarily travel throughout the unburnt gas at uniform speed relative to the wall of the tube, for the gas may be moving at varying velocities at different distances from the flame. Waves travelling from hot to cold gases diminish in speed, whilst they increase in speed on travelling from cold to hot gases; on meeting a flame they are partially reflected, with change of sign under suitable conditions, and partially transmitted. On collision with a solid obstruction they are reflected, usually at a lower speed.

Detonation may be set up either ahead of or, more usually, within the flame front, due to the effect of (1) waves travelling in front of or from behind the flame, (2) the collision or overtaking of wave and flame or wave and wave, or (3) the collision of a wave with an obstruction or the closed end of the tube.

The Adsorption of Hydrogen on Tungsten

By J. K. Roberts, Laboratory of Colloid Science, Cambridge

(*Communicated by Eric K. Rideal, F.R.S.—Received June* 5—*Revised July* 17, 1935)

1—Introduction

When a gas atom or molecule strikes a solid surface it may either rebound, the collision being accompanied by an interchange of energy, or it may be adsorbed. In order to obtain a satisfactory interpretation of either type of phenomenon it is necessary to know the nature of the atoms concerned. This is comparatively easy for the gas, but the solid requires closer attention. The essential condition is that the surface of the solid is completely freed from adsorbed films of impurity before the experiment starts and remains free.

In the earlier papers* in which an account of the development of the present work on this problem has been given, it has been shown that the average interchange of energy when gas atoms strike and rebound from a slightly hotter metal surface is very markedly affected by the presence of such adsorbed films. The accommodation coefficient is a quantitative measure of the efficiency of this interchange and for neon with a tungsten surface the removal of all adsorbed films caused the accommodation coefficient to change almost by an order of magnitude (0·6 for an ordinary surface covered with adsorbed films of impurity and about 0·07 or 0·08 for a bare tungsten surface). This suggested that the accommodation coefficient of neon could be used as an indicator in studying the adsorption on bare tungsten of known gases mixed with the neon. The application to the study of the adsorption of hydrogen of the method that has been developed with this idea as a basis is described* in § 2.

---

* Roberts, 'Proc. Roy. Soc.,' A, vol. 129, p. 146 (1930); vol. 135, p. 192 (1932); vol. 142, p. 518 (1933).

A second new experimental method has been used in which it is possible under certain conditions to measure heats of adsorption on a single fine wire. This is described in § 3. The methods are being used with all the metal-gas systems to which they are applicable.

2—ACCOMMODATION COEFFICIENT OF NEON

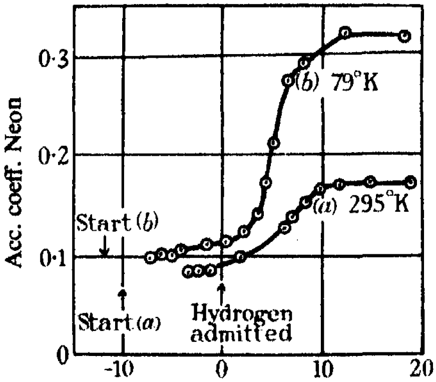
The apparatus for measuring accommodation coefficients, described in one of the earlier papers,† was modified by the addition of a side tube leading to a gas pipette, by means of which small amounts of hydrogen could be admitted at will to the neon, which was continuously circulated through suitably placed charcoal tubes immersed in liquid air, so that it was kept free from any adsorbable impurities coming off the glass. The hydrogen was introduced into a reservoir connected to the pipette by heating a palladium tube.

As in the earlier experiments, the neon, before the admission of any hydrogen, was circulated for some time, and the wire was flashed at a temperature well above 2000° K to remove any adsorbed films from its surface. When the wire had cooled down, a measured current was passed through it sufficient to raise its temperature ten or twenty degrees above that of the bath in which the containing tube was immersed, and readings of resistance and time were taken, the pressure of the neon being measured on a MacLeod gauge. From each resistance measurement a value of the accommodation coefficient was deduced. These values showed a slow and, in some experiments, almost inappreciable drift with time, showing that the amount of adsorbable impurity in the neon was extremely small.* After a short time a trace of hydrogen was admitted to the stream of neon. The accommodation coefficient began to increase and rose comparatively rapidly to a steady final value of 0·17 at 295° K and 0·32 at 79° K, as shown in fig. 1, in which the accommodation coefficient of neon is plotted as a function of the time, curve (*a*) referring to 295° K and curve (*b*) to 79° K. The time zero is taken as the time at which the hydrogen was admitted, and is indicated by the arrow. The time at which each experiment started, that is when the flashing current was cut off, is also indicated by an arrow. The results given refer to an experiment in which the pressure of the neon was 0·07 mm and that of the hydrogen introduced was 4×10^{-4} mm, the latter pressure being obtained by admitting the same amount of hydrogen into the evacuated apparatus, and measuring the pressure produced on the MacLeod gauge.

---

* A short preliminary account of these experiments has been given in a paper in the Proc. Camb. Phil. Soc.,' vol. 30, p. 74 (1934). This paper contains an account of the experimental method and a short discussion of the nature of the adsorption process. As a fuller discussion is now possible in the light of further results that have een obtained, and as the method has also been applied to the study of oxygen films, it will be convenient to include here a brief account of the experimental technique.

† 'Proc. Roy. Soc.,' A, vol. 135, p. 192 (1932).



Time in Minutes

FIG. 1.

---

* It may be mentioned that the phenomena about to be described would be completely masked if this residual drift were not very small. Mann, 'Proc. Roy. Soc.,' A, vol. 146, p. 779 (1934), appears to imply a criticism of the linear extrapolation to zero time used in the earlier papers to obtain the accommodation coefficient of a bare surface. Where the drift due to contamination is very small this linear extrapolation is fully justified. In some of the first experiments carried out the purity was not so high as later, and these showed a more marked drift and not a straight line. The curves in such cases were extrapolated to zero time by fitting a suitable exponential formula to them, and the extrapolated values for a bare surface did, in fact, agree with those obtained in the later and better experiments when the drift was negligibly small.

The change in accommodation coefficient must be due to the adsorption of hydrogen on the tungsten surface at this low partial pressure. To show that it was not due to small traces of oxygen in the hydrogen, the experiment was carried out admitting, instead of hydrogen, the same volume of air at the same partial pressure as the hydrogen in the above experiment. No trace of the phenomenon described above was observed; that is, the charcoal tubes removed completely this relatively very large amount of oxygen. The possibility that the effect is due to traces of oxygen is therefore completely excluded.

Further experiments showed that the final value of the accommodation coefficient of the neon was not affected appreciably if the partial pressure of the hydrogen was increased to 3×10^{-3} mm—that is nearly ten times.* This indicates that the state of the adsorbed film is not affected by increasing the pressure of the gas nearly ten times. We are thus forced to the conclusion that the film is saturated at the lower pressure.†

A further characteristic is that the adsorption process is not a slow one, saturation being reached in a short time even at these low pressures and at 79° K.

3—MEASUREMENT OF HEAT OF ADSORPTION

Since saturation probably means the formation of a complete monomolecular film, the fact that this occurs at very low pressures of hydrogen, and that the adsorption takes place rapidly, raises the possibility of measuring the heat of adsorption on a single fine wire. It is, of course, necessary to use a wire or fine strip if the surface is to be bare before the hydrogen is admitted to it, and only with a bare surface is this type of adsorption observed.

---

* At these higher partial pressures it was necessary to make a correction for the heat carried away from the wire by the hydrogen, since hydrogen, owing to its smaller mass and to the fact that it has rotational as well as translational energy, is a more efficient agent than neon at the same pressure. In making the correction it was assumed that the accommodation coefficient of hydrogen on a tungsten surface was 0·22. This is the "*initial*" value given by Blodgett and Langmuir ('Phys. Rev.,' vol. 40, p. 78 (1932)), and refers to tungsten flashed *in vacuo* to which hydrogen is then introduced.

† As a result of a certain line of reasoning, Blodgett and Langmuir (*loc. cit*, p. 104, a portion of the last sentence on the page) say: ". . . we believe that we never have a bare W surface in contact with hydrogen at low temperatures, but instead of this an HW surface. . . ." The present work provides direct experimental evidence that this view is correct even at very much lower pressures than were considered by them. Direct evidence for such a view must be of the nature of that given in the present paper, that is, the difference between the bare and the covered surface must be observed directly.

The idea underlying the method was that the wire itself on which the adsorption took place should form the calorimeter. There is a great advantage in having the heat liberated just where it is wanted, and in knowing the apparent area of the surface on which adsorption takes place. The thermal capacity of the wire is known from its diameter (0·0066 cm) and length (28·2 cm), and the density and specific heat of the material of which it is made. If its mean rise of temperature when the adsorption takes place can be measured, the total heat given out by the adsorption process can be obtained immediately. The measurement of the amount of gas adsorbed will be considered in due course.

The apparatus* is shown in fig. 2. The wire was contained in a tube W protected from mercury and other vapours by liquid air traps. This tube and the tube G containing the Pirani gauge were immersed in a large Dewar vessel filled with oil. The necessary small amount of hydrogen was obtained by expanding one charge of the gas pipette P connected to the hydrogen reservoir into a large volume B (not shown), and then admitting to the part of the apparatus containing the wire the amount of this expanded gas contained in the small volume A between the two mercury cut-offs.

The actual experimental procedure in its simplest form was as follows:

(i) The wire was flashed and when the flashing current was cut off the wire was connected to a sensitive and carefully designed bridge, in which a Paschen galvanometer was used. As the wire cooled down the galvanometer showed a drift with time. After about ten minutes this drift was sufficiently slow to be easily followed and controlled by altering one of the other arms of the bridge. When this stage was reached the apparatus was cut off from the pumps, the hydrogen charge was prepared, and regular readings of the galvanometer deflection and time were taken. A small amount of hydrogen was then admitted to the apparatus, and the galvanometer showed a deflection in the direction corresponding to a rise of temperature of the filament as shown in fig. 3. The time of admission, the maximum reading, and the time of its occurrence were noted. The

---

* The method was first tried out with the assistance of B. Whipp, and a short account of these preliminary experiments was published (' Proc. Camb. Phil. Soc.,' vol. 30, p. 376 (1934)) when he left the department. The purpose of this preliminary work was to show that the method is a practicable one. To do this it was necessary to establish in a general way that the measured heat of adsorption was independent of various factors, and in particular of the following: (*a*) The amount of gas adsorbed within reasonable limits; in the later experiments described above it has been possible to trace the variation of heat of adsorption as the surface becomes more and more covered; (*b*) the bridge current used; (*c*) the amount of gas remaining after adsorption which determines the final heat loss from the wire. It was also shown by storing the hydrogen over charcoal in a tube immersed in liquid air that the effects observed were not due to traces of oxygen in the hydrogen. To show that hydrogen was not being adsorbed on tungsten evaporated from the filament on to the glass, the flashing current was varied enough to alter the vapour pressure of the tungsten by a large factor, but there was no effect on the heat of adsorption. The significance of these points will be realized when the details of the method have been given, and they need not be discussed further. The method has since been developed, using the new and improved apparatus described above.

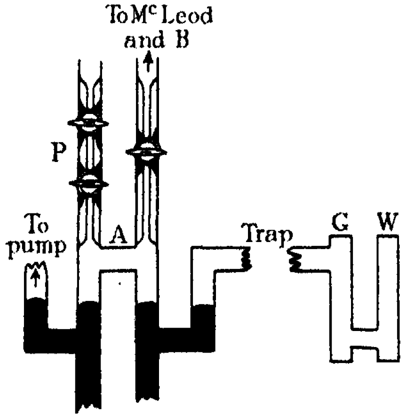


Fig. 2.

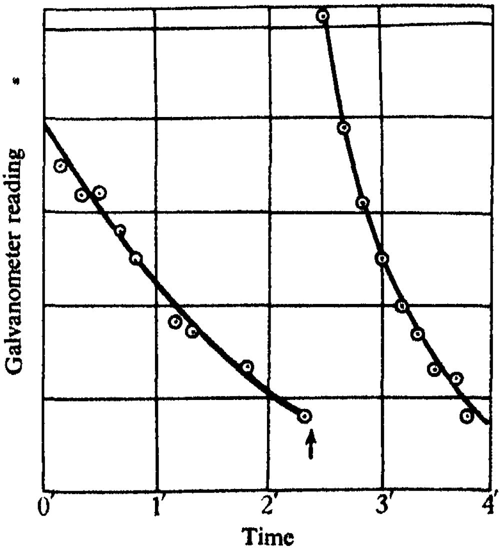


Fig. 3.

deflection of the Pirani gauge was read at the same time. The hydrogen was now pumped out, the deflection of the Pirani gauge on evacuation being measured as a check.

(ii) The volume A was then refilled from B so that the pressure in it was the same as in (i).* This hydrogen was admitted to the evacuated apparatus without flashing the wire. The deflection of the Pirani gauge was measured, and was greater than for the same amount of hydrogen in (i) because there was no adsorption on the unflashed wire.† At the same time it was observed, as would be expected, that there was practically no deflection of the Paschen galvanometer. This hydrogen was in general pumped out and the Pirani gauge deflection again read, but in some experiments, as a check, a second lot was admitted without pumping out the first lot and the Pirani gauge deflection measured. The Pirani gauge was calibrated by diminishing the bridge sensitivity in a measured ratio and comparing the gauge directly with a MacLeod gauge.

Provided the volume of the apparatus is known, the Pirani gauge readings give directly the amount of gas adsorbed on the wire in (i). The volume was determined in the ordinary way by expanding air at a few cm pressure from a measured volume into the apparatus, and measuring the pressure change.‡

To obtain the temperature rise of the filament from the readings a correction had to be made for the cooling during the time when the deflection was occurring. This correction, which was small and never more than about 10% of the total deflection, was obtained in the usual way by a study of the slope of the temperature time curve at different points. The change of resistance corresponding to a given galvanometer deflection was obtained by altering by a small known amount a large shunt in the opposite arm of the bridge, the resistance of which was equal to that of the filament. The temperature change was deduced from the temperature coefficient of tungsten from the same reel which had been measured in the earlier accommodation coefficient experiments. The heat evolved in the adsorption process is given by the expression $m\ s\ \Delta T$, where m is the mass of the wire, s the specific heat, and ΔT the rise of temperature.

---

* Since B was not infinite these two pressures were not identical. A correction of about 1% determined from the volumes of A and B was made for this.

† The method is applicable only to an adsorption process which is practically complete at very low pressures, for only in such a case is the amount adsorbed on the wire comparable with the total amount of gas in the apparatus, and the thermal disturbance produced by the residual gas negligible.

‡ In obtaining the number of moles adsorbed it was necessary to take into account the fact that the density of the gas in the liquid air traps was different from that in the rest of the apparatus, and, since at the low pressures used the free path of the gas molecules is large compared with the diameters of the tubes, in making the calculation of the effect of this density difference, to remember that under these conditions the equilibrium condition is $p_1/p_2 = (T_1/T_2)^{\frac{1}{2}}$, where p_1 and p_2 are the respective pressures in parts of the apparatus at temperatures T_1 and T_2.

The above description gives the principle of an experiment which determines the average heat of adsorption with the surface at the end fully covered. This procedure was not followed exactly, but the amount of hydrogen admitted in one charge of A was made less than was necessary to cover the surface completely. When the first charge was admitted there was a heating effect, but there was no pressure change sufficient to be indicated on the Pirani gauge (that is, it was less than 2 or 3×10^{-7} mm). Thus even at these low pressures adsorption was complete and rapid.* Charge after charge was admitted, and the heating effect measured each time. Finally, when saturation was reached, the Pirani gauge showed a deflection, and from this it was possible to deduce the amount admitted at each charge and so to obtain the variation of heat of adsorption as the surface became progressively more covered.

The results of these experiments are collected together in Table I, and are discussed in the following sections.

4—NUMBER OF MOLECULES ADSORBED

It is of interest to compare the total number of hydrogen molecules adsorbed with the number of tungsten atoms in the surface of the wire. In the (110) plane of tungsten† there are $14 \cdot 24 \, . \, 10^{14}$ atoms per cm². The superficial area of the filament was $0 \cdot 58$ cm², which would give $8 \cdot 3 \times 10^{14}$ tungsten atoms in the surface. This must be reduced by a small factor on account of the fact that when the wire was flashed short lengths near the ends were below 2000° K, so that these lengths would probably not be completely freed from oxygen. Calculation showed that this can be allowed for by multiplying by $26 \cdot 6/28 \cdot 2$, giving effectively $7 \cdot 8 \times 10^{14}$ tungsten atoms in the surface if it were smooth.

We shall see, as the results are discussed, that a satisfactory and consistent quantitative account of the whole process can be given if the

---

* It is not possible to deduce from these results a value for the possible energy of activation of the adsorption process which would really have any significance. The essential point is that it is, in fact, effectively an instantaneous process. For a short discussion, *see* the footnote Roberts, 'Proc. Camb. Phil. Soc.,' vol. 30, p. 76 (1934).

† The reason for using the (110) plane has been given by Langmuir, 'J. Amer. Chem. Soc.,' vol. 54, p. 2819 (1932).

TABLE I

Experiment	Admission	Molecules H_2 adsorbed	Mean number unfilled spaces on surface	Heat of adsorption *k.* calories per mol H_2	Symbol used in plotting
I	1	1·19 × 10^{14}	7·18 × 10^{14}	34·2	⊙
	2	1·18	4·80	28·5	
	3	1·17	2·46	20·3	
	4	0·64	0·64	17·5	
		Total = 4·18			
II	1	0·99	7·84	—	⊡
	2	0·98	5·86	34·2	
	3	0·97	3·92	31·1	
	4	0·96	1·98	25·3	
	5	0·51	0·52	22·7	
		Total = 4·41			
III	1	0·99	7·68	38·6	●
	2	0·98	5·70	36·7	
	3	0·97	3·76	—	
	4	0·96	1·82	21·5	
	5	0·43	0·44	—	
		Total = 4·33			
IV	1	1·01	5·68	40·3	+
	2	1·00	3·66	31·1	
	3	0·99	1·68	22·8	
	4	0·34	0·34	—	
		Total = 3·34			
V	1	1·16	6·44	—	×
	2	1·15	4·14	24·8	
	3	1·14	1·84	21·2	
	4	0·35	0·36	—	
		Total = 3·80			

hydrogen is adsorbed as atoms. If the wire were smooth and if there were one atom of hydrogen per atom of tungsten in the surface, the number of molecules that could be adsorbed would be $3 \cdot 9 \times 10^{14}$.

In experiments I, II, and III the observed numbers adsorbed were $4 \cdot 2$, $4 \cdot 4$, and $4 \cdot 3 \times 10^{14}$ respectively. In order to reconcile these with the number of spaces in the smooth surface, it must be supposed that the roughness of the wire increases its effective area in the ratio $4 \cdot 4/3 \cdot 9$, or $1 \cdot 1$ times. This is reasonable, and may be compared with $1 \cdot 4$ deduced by Taylor and Langmuir* from experiments on the adsorption of caesium on tungsten. We conclude, therefore, that each tungsten atom on the surface adsorbs one atom of hydrogen.

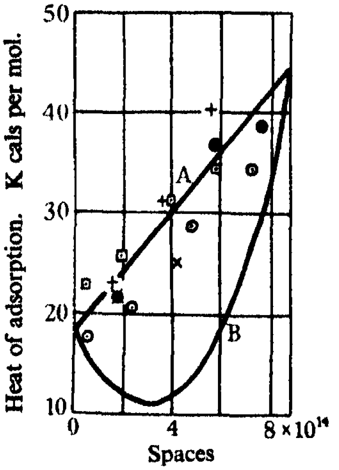


Fig. 4.

5—Heat of Adsorption

The heats of adsorption of successive lots of hydrogen admitted to the filament flashed at the beginning of each experiment, but not in between the admissions numbered 1, 2, 3, etc., are given in the last column of Table I and are plotted as a function of the number of unfilled spaces† on the surface in fig. 4. The relative values of the heats for the successive admissions in a particular experiment can be fixed with a very much higher accuracy than the absolute values which are used in comparing different experiments.

To deduce the heat of adsorption when a hydrogen molecule strikes a bare surface, we extrapolate to the value corresponding to $8 \cdot 8 \times 10^{14}$, the maximum number of atoms adsorbed, and obtain 45 kilocalories per mol H_2. On the other hand, if a hydrogen molecule strikes a place on the surface where two adjacent atoms are bare and all the surrounding places within appreciable range are already covered, the heat of adsorption is 18 kilocalories per mol H_2.

---

* Taylor and Langmuir, 'Phys. Rev.,' vol. 44, p. 423 (1933).

† The method used to calculate the number of unfilled spaces can be explained by considering experiment I in detail. After admission 4 no further adsorption occurs, so there are no unfilled spaces. During admission 4 there are $0 \cdot 64 \times 10^{14}$ molecules, *i.e.*, $1 \cdot 28 \times 10^{14}$ atoms, adsorbed. The mean number of unfilled spaces for admission 4 is $(1 \cdot 28/2) \times 10^{14}$, *i.e.*, $0 \cdot 64 \times 10^{14}$. At the end of admission 3 there are $1 \cdot 28 \times 10^{14}$ unfilled spaces, and during admission 3 there are $3 \cdot 34 \times 10^{14}$ atoms adsorbed. The mean number of unfilled spaces for admission 3 is $\{1 \cdot 28 + (3 \cdot 34/2)$ $\times 10^{14}$, *i.e.*, $2 \cdot 46 \times 10^{14}$. The calculation is carried out in a similar way for admission 1 and 2. In this method of calculation it is assumed that there is no contamination of the wire between admissions, and variation in the total number of molecules adsorbed in different experiments would be attributed to an initial small contamination of the surface.

6—THEORY OF HEAT OF ADSORPTION. ELECTRICAL PROPERTIES OF HYDROGEN FILM

It has already been mentioned that there is strong evidence for the view that the hydrogen is adsorbed as atoms with one atom for each tungsten atom in the surface. We shall assume as before that we have to do with a (110) plane of the tungsten crystal. In this the atoms are arranged at the intersections of a rectangular mesh, the sides of the elementary rectangles being a and $a\sqrt{2}$, and there is one further atom at the centre of each rectangle, so that each atom is surrounded by four equally removed others at a distance $a\sqrt{3}/2$.

Now when, at a given stage in the adsorption process, a molecule is adsorbed on two adjacent bare tungsten atoms, a certain number of the other three places immediately surrounding each of the spaces where one of the atoms is adsorbed will be occupied. For an immobile film, in which every molecule striking a possible pair of vacant spaces on the surface is adsorbed, this number is on the average strictly proportional to the total number of occupied places on the surface.

Let the energy in calories per gram-atom required to remove one hydrogen atom be as follows: If there is only one hydrogen atom on the surface, Q. If the four places surrounding the atom in question are occupied by hydrogen atoms and the rest of the surface is bare, $(Q - 4X_1)$. If the whole surface is covered, $(Q - 4X_1 - X_2)$, the term $4X_1$ arising from the four nearest atoms and the term X_2 from the rest. For the sake of clearness it is convenient to assume at first in this way that the effect of an adsorbed atom at a given distance is independent of the fraction of the surface covered. This assumption can easily be removed later.

Let us obtain expressions for the heat of adsorption of a molecule of H_2 into an immobile film under various conditions, if D calories per mol H_2 is the heat of dissociation in the gas state.

(*a*) *On a bare surface*—

First step—Dissociate the molecule into atoms. Heat evolved $= -$ D calories per mol H_2.

Second step—Adsorb atom 1 on the surface. Heat evolved = Q calories per gram-atom.

Third step—Adsorb atom 2 on a space next to that occupied by atom 1. Heat evolved = $(Q - X_1)$ calories per gram-atom.

Total heat evolved $= 2Q - D - X_1 = 45{,}000$ calories per mol H_2. (1)

(*b*) *On a surface completely covered, except for the two neighbouring spaces on which the adsorption takes place—*

First step—Dissociate the molecule into atoms. Heat evolved $= -D$ calories per mol H_2.

Second step—Adsorb atom 1 on the surface. Heat evolved $= (Q - 3X_1 - X_2)$ calories per gram-atom.

Third step—Adsorb atom 2 on the surface. Heat evolved $= (Q - 4X_1 - X_2)$ calories per gram-atom.

Total heat evolved

$$= 2Q - D - 7X_1 - 2X_2 = 18{,}000 \text{ calories per mol } H_2. \qquad (2)$$

(*c*) *The average value on two neighbouring spaces on a surface a fraction* θ *of which is already covered—*

In this case we obtain similarly—

Total heat evolved

$$= 2Q - D - 2\theta\,(3X_1 + X_2) - X_1 \text{ calories per mol } H_2. \qquad (3)$$

Combining equations (1) and (2), we obtain

$$3X_1 + X_2 = 13{,}500 \text{ calories}. \qquad (4)$$

This, of course, depends on the assumptions made that the effect of an atom at a given distance does not depend on the fraction of the surface covered, and that the film is immobile.

The relative magnitudes of X_1 and X_2 can be determined only if the nature of the forces which give rise to them is known. It will now be assumed that the forces are electrostatic. In considering the equation of state of caesium atoms adsorbed on tungsten, Langmuir* has treated the question of such electrostatic forces in detail, and has pointed out that, if the separation of charge between the adsorbed atom and the surface produces for each atom a dipole of moment M e.s.u., the potential energy of two adsorbed atoms a distance r cm apart is $M^2/2r^3$ ergs, *i.e.*, one-half that between two dipoles. We consider one adsorbed atom and calculate the potential energy of this atom due to all the other adsorbed atoms on a completely covered surface.

---

* Langmuir, 'J. Amer. Chem. Soc.,' vol. 54, p. 2816 (1932).

We have just seen that, when we are concerned with the adsorption as atoms of diatomic molecules, the four nearest atoms must be considered separately from the rest. In accordance with the notation already used, let x_1 ergs be the contribution of each of these four atoms to the total potential energy, where

$$x_1 = (X_1/N_0) \,.\, 4 \cdot 18 \,.\, 10^7. \tag{5}$$

Similarly, let x_2 ergs be the contribution of all the other atoms to the potential energy. The distribution of the atoms has been discussed at the beginning of this section. To calculate the total potential energy, a circle of radius $3 \cdot 2a$ was drawn around the atom under consideration, and the contribution of each atom inside this circle was calculated individually, while that of the atoms outside was calculated by integration, assuming a uniform distribution of particles of surface density $\sigma = \sqrt{2}/a^2$ per unit area. In this way three-quarters of the total value of x_2 was calculated directly, and only the remaining quarter by integration. It was found that

$$\left.\begin{aligned} x_1 &= 0 \cdot 77 \ M^2/a^3, \\ x_2 &= 5 \cdot 12 \ M^2/a^3. \end{aligned}\right\} \tag{6}$$

It may be mentioned here that the change in the work function of the surface when an adsorbed film is formed is

$$2\pi \ M \ \sigma \ \text{e.s.u.}, \tag{7}$$

where σ is the surface density of adsorbed atoms.

There is experimental evidence that for caesium films on tungsten* the electric moment per adsorbed atom diminishes as the surface becomes more and more covered. Langmuir has explained this as being due to the depolarizing effect on a given adsorbed atom of the field due to all the other adsorbed atoms. Taylor and Langmuir (*loc. cit.*, fig. 14, p. 437) give for caesium films V_c, the contact potential against bare tungsten, as a function of θ, the fraction of surface covered. V_c/θ is proportional to M, and, if from Taylor and Langmuir's data we deduce the value of V_c/θ and plot it against θ, we obtain a nearly linear curve which can be extrapolated to $\theta = 0$, and to $\theta = 1$, giving $V_c/\theta = 9 \cdot 8$ for $\theta = 0$, and $V_c/\theta = 2 \cdot 6$ for $\theta = 1$. The ratio of these is $3 \cdot 8$, so the variation in this case is considerable.

---

* Taylor and Langmuir, 'Phys. Rev.,' vol. 44, p. 423 (1933). Compare also for thorium films, Langmuir, 'Acta phys. chim. U.S.S.R.,' vol. 1, p. 378 (1934).

If M varies in a similar way for the hydrogen film, the effect from our point of view is important, as it modifies equation (3) and makes the relation between heat of adsorption and θ non-linear. To show this, let us assume a linear variation of M with θ and put

$$M = M_0 \left(1 - \frac{M_0 - M_1}{M_0}\,\theta\right), \tag{8}$$

where M_0 is the value of M for $\theta = 0$, M_1 for $\theta = 1$. X_1 and X_2 in equation (3) will not now be constants, but will vary with θ and to a first approximation* we may take both proportional to M^2; that is

$$\left.\begin{aligned} X_1 &= 0 \cdot 77 f M^2 \\ X_2 &= 5 \cdot 12 f M^2 \end{aligned}\right\}, \tag{9}$$

where the factor f includes the a^3 factor and converts to appropriate units. Substituting from equations (8) and (9) in (3), and assuming as a definite example that $M_0 = 2\,M_1$, we obtain for the heat of adsorption per mol H_2

$$2Q - D - 3 \cdot 08 f M_1^2 - 4\,(14 \cdot 09\,\theta - 14 \cdot 67\,\theta^2 + 3 \cdot 72\,\theta^3) f M_1^2. \tag{10}$$

We know (*see* fig. 4) that for $\theta = 0$ the heat of adsorption is 45,000 and for $\theta = 1$ it is 18,000. Using these results in equation (10), we obtain the curve marked B in fig. 4. If, on the other hand, M = constant, we should obtain the straight line A.

It is quite clear that the curve B does not fit the observations at all.† We are therefore forced to the conclusion that M is almost independent of the fraction of surface covered. A detailed study of the electrical properties of the hydrogen film will make a fuller interpretation possible. Such an investigation will be carried out shortly.

From equation (2) we obtain

$$Q - 4X_1 - X_2 = (D + 18)/2 - X_1/2. \tag{11}$$

This gives the heat per gram-atom required to remove the first atom of hydrogen from a completely covered surface. From (5), (6), and (4) we obtain

$$X_1 = 1400 \text{ calories}. \tag{12}$$

---

* This neglects any disturbance produced in its neighbourhood by the molecule that is being adsorbed. The effect of such disturbance will be greater on X_1 than on X_2.

† When an atom is adsorbed, the dipole moments of the surrounding adsorbed atoms are reduced. The mutual potential energy of these atoms is thus altered. An approximation calculation of the effect of including the additional term due to this shows that the fit with the observations is made worse. An approximate calculation for a mobile film with similar variation of M indicates that the departure from linearity is also very marked.

The numerical results which can be deduced from the experiments using these equations are given in Table II. D is taken[*] as 101 kilocalories per mol H_2 and a as $3 \cdot 1 \times 10^{-8}$ cm.

TABLE II

Heat to remove first atom from complete film *k*. calories per gram-atom	Heat to remove last atom from surface *k*. calories per gram-atom	Dipole moment per adsorbed atom e.s.u.	Change in work function by covering surface complete film volts
58·8	73·7	$1 \cdot 94 \times 10^{-18}$	5·2

It has been mentioned above that the contact potential of a tungsten surface completely covered with caesium atoms ($3 \cdot 6 \times 10^{14}$ per cm^2) against a bare tungsten surface is 2·6 volts. This gives $3 \cdot 9 \times 10^{-18}$ e.s.u. for the moment per adsorbed caesium atom. The order of magnitude is the same as for the hydrogen film, but for the latter the effective separation of the charges is less. The sign of the dipole is similar to that of an oxygen film, *i.e.*, opposite to that of a caesium film.[†]

7—STABILITY OF THE ADSORBED FILM

To obtain the temperature at which the adsorbed film becomes unstable, the wire was first covered with hydrogen, and then heated with a measured current i for one minute with the apparatus open to the pumps. When, after cutting this current off, the temperature had become sufficiently steady, a charge of hydrogen was admitted as before, and observations of the heating effect on the filament were taken.

As i was increased a current i_0 was reached at which the heating effect due to adsorption (*i.e.*, the amount of adsorption) became appreciable, and began to increase rapidly with increase of i. The temperature T_0 at the middle of the filament when the current is i_0 is the temperature at which the film is unstable for times of the order of one minute. T_0 was obtained from Langmuir's tables for ideal tungsten filaments.[‡] The results obtained in this way are shown in fig. 5, in which the amount of adsorption in arbitrary units is plotted as a function of the maximum (central) temperature to which the filament has been heated for one minute. It will be seen that the rise in the curve occurs at about 700° K or just below.

---

[*] *See* Hinshelwood and Williamson, " Reaction between Hydrogen and Oxygen," p. 26 (Oxford, 1934).

[†] This volume, p. 479.

[‡] A calculation using the data given by Langmuir, MacLane, and Blodgett (' Phys. Rev.,' vol. 35, p. 478 (1930)) showed that for the wire used (a hairpin filament 0·0066 cm diameter and total length 28·2 cm), the effect of end losses on the middle temperature was unimportant if the temperature of the middle was greater than about 600 to 650° K. This is just below the important range, so that no correction for end losses was necessary.

In another paper* the evaporation of adsorbed atoms or molecules from surfaces is considered, and the formula

$$t_f = -2{\cdot}06 \, . \, 10^{-13} \, . \log_{10} f \, . \, (M^{\frac{1}{2}}T^{\frac{1}{2}}/\,\Phi) \, . \, e^{\Phi/RT} \tag{13}$$

is obtained, where t_f is the time in seconds taken for a film to evaporate to such an extent that a fraction f of the initial number of spaces occupied remains occupied, M is the molecular weight of the evaporating particles ($O_2 = 32$), Φ is the heat of desorption in calories per mol which is assumed constant, and the numerical constant is determined from the experiments of Langmuir and Villars on the evaporation of oxygen.

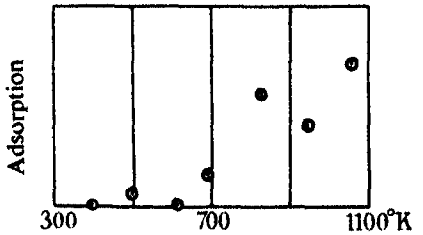


Fig. 5.

We have seen in Table II that the present measurement of the heat of adsorption shows that for the desorption of hydrogen atoms from a complete film on tungsten, $\Phi = 58{,}800$ calories per gram-atom. The values of $\log_{10} t_{\frac{1}{2}}$ calculated from equation (13) for $\Phi = 57$ and $\Phi = 60$ kilocalories per gram-atom are shown in fig. 6 plotted against $1/T$. We want to find the temperature at which it would be expected according to equation (13) that the film would become unstable for times of the order of one minute. This will occur when $\log_{10} t_{\frac{1}{2}}$† is of the order 2.

It will be seen that the formula indicates that the film should begin to be unstable at temperatures of about 700° K. This is in very remarkable agreement with the experimental result, considering that the formula contains no adjustable constant. This can hardly be a coincidence, and we are justified in concluding that the assumptions on which the calculation is based are correct; that is, that the adsorbed film is an atomic one, and that the evaporation formula is a good first approximation.

---

* Roberts, 'Trans. Faraday Soc.,' in course of publication.

† The actual value of f chosen is not important: for example, if we calculate for $f = 4/5$ instead of $f = 1/2$, the times are multiplied by one-third. It may be noted also that variation of Φ as the atoms evaporate is not of great importance, as, if the wire is maintained at temperature T for one minute, the calculations will indicate the value of T at which the film will begin to break up.

For the last atoms leaving the surface it is seen from Table II that $\Phi = 73$ kilocalories per gram-atom. Using this, the evaporation rates have been calculated from (13), and are also shown in fig. 6. The calculations indicate that to remove the last traces of hydrogen from the surface it is necessary to heat it to 900° K, or a little higher, for about one minute.

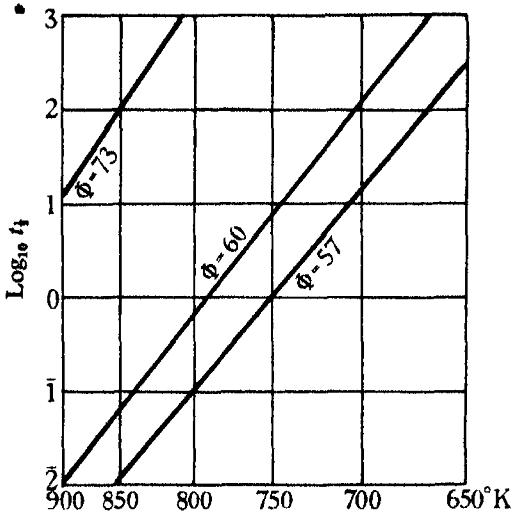


FIG. 6.

8—BEARING OF THE RESULTS ON THE GENERAL PROBLEM OF THE ADSORPTION OF HYDROGEN

The results are significant in connexion with activated adsorption. Roughly speaking, the characteristics of this are (*a*) that over a certain temperature range at a given pressure the amount of gas taken up increases with the temperature, and (*b*) that the adsorption process is a slow one, sometimes extremely slow, taking hours or even days to reach saturation at pressures of the order of centimetres of mercury. The phenomenon has been shown to occur with hydrogen on a number of surfaces: for example, nickel, tungsten, iron, copper and various mixed oxides. The behaviour with tungsten* is quite characteristic. Since the present experiments have been carried out with tungsten, this is important, as it makes it probable that what is said applies to all metals.

Most attempts made to explain the occurrence of this type of adsorption on metals involve the two following general assumptions: (i) that the

---

* Frankenburger and Hodler, 'Trans. Faraday Soc.,' vol. 28, p. 229 (1932).

system involved is the system metal-hydrogen, (ii) that the process is one of chemi-sorption as opposed to van der Waals, or physical adsorption.

Assumption (i) is generally made tacitly, but the present experiments show that it is inadmissible, since the behaviour of a bare surface is quite different from that observed in the type of experiment on which the idea of activated adsorption is based. It must be concluded that in these other experiments the surface was not a bare metal surface but was, before the hydrogen was admitted, at least partially covered with a layer of adsorbed gas, and that the system involved was this complicated one and not the simple system metal-hydrogen. The possible importance of surface impurities in connexion with the problem of activated adsorption has been pointed out by Burrage,* who has suggested that there is internal evidence in the experiments on activated adsorption themselves that such impurities are playing an important part in the processes that are occurring. The position, however, only becomes really clarified when observations have been made on a surface which is bare, so that the behaviour of such a surface is known. We shall discuss further, from this point of view, the behaviour of surfaces already covered with an adsorbed film and oxide surfaces in another paper.†

With regard to assumption (ii), the present experiments show that in the case of tungsten the process of activated adsorption is not the simple one of chemi-sorption of hydrogen on tungsten; for, when hydrogen actually comes into contact with tungsten, the surface is completely covered with a chemi-sorbed film in a few seconds or less at room temperature, the pressure of the hydrogen being 10^{-4} mm or less. Furthermore, even at − 190° C and at pressures of this low order, the film is formed as far as the apparatus (accommodation coefficient experiments) can detect as rapidly as at room temperature. Nothing could be less like the phenomena associated with the term activated adsorption than this. It is necessary to point this out specifically because the terms chemi-sorption and activated adsorption are sometimes used as if they are synonymous.‡ Furthermore, elaborate theories have been proposed to show how and why chemi-sorption of hydrogen on metals takes place slowly. Such theories can only be maintained if they can be modified to show that chemi-sorption should take place rapidly.

---

* Burrage, 'Trans. Faraday Soc.,' vol. 29, p. 677 (1933). Other references will be found in this paper.

† This volume, p. 480.

‡ As an example the following from a paper published in December, 1934, may be quoted: "The chemi-sorption (activated adsorption) of the hydrogen. . . ."

It will be realized, too, that, at any rate down to liquid air temperatures, van der Waals adsorption of hydrogen can be observed only after the metal surface is covered with an adsorbed film of some sort.

The first experiments on the adsorption of hydrogen were carried out when I held a Moseley Studentship at the Cavendish Laboratory, and I should like to thank Lord Rutherford and the Council of the Royal Society for their kindness at this time. I should also like to thank Professor Rideal for his encouragement during the time that the experiments have been continued in the Department of Colloid Science.

SUMMARY

A method has been developed for studying the adsorption of hydrogen on a bare tungsten surface which depends on the fact that the accommodation coefficient of neon is different for a bare surface and for a surface with an adsorbed film on it. It has been shown that saturation occurs at a partial pressure of hydrogen of the order of 10^{-4} mm or lower (*see* below), so that the process is one of chemi-sorption. This chemi-sorption takes place rapidly at these low pressures and at 79° K.

The heat of adsorption and amount of gas adsorbed have been measured directly on a single fine wire and the variation of heat of adsorption traced with variation of the covering of the surface. The film has been shown to be stable at negligibly low pressures at room temperature. The variation with temperature of the stability of the film has also been studied, and thus, using the measured heat of adsorption and a new desorption formula, the film has been shown to be atomic. There is one atom of hydrogen per atom of tungsten in the surface. From the results deductions have been drawn concerning the electrostatic forces between the adsorbed atoms and the change in contact potential when a bare surface is covered with a complete hydrogen film. The binding is of the same type as that of oxygen on tungsten. The bearing of the results on the general problem of hydrogen adsorption and, in particular, on the question of activated adsorption is discussed. Values for the accommodation coefficient of neon on a tungsten surface covered with hydrogen are given for 295° and 79° K.

Some Properties of Adsorbed Films of Oxygen on Tungsten

By J. K. Roberts, Laboratory of Colloid Science, Cambridge

(*Communicated by E. K. Rideal, F.R.S.—Received June 5—Revised July* 17, 1935)

In the foregoing paper methods have been described for studying the adsorption of hydrogen on tungsten. Similar experiments have been carried out with oxygen. It is known from thermionic work that oxygen forms an adsorbed film on tungsten which is stable up to nearly 2000° K. The present experiments show that on top of this a second film is formed which is quite stable at ordinary temperatures. The properties of both films have been investigated.

1—Accommodation Coefficient of Neon

We shall first consider the results obtained, using the accommodation coefficient of neon as an indicator. The experiment was carried out in the same way as with hydrogen except that the oxygen (prepared by heating potassium permanganate) was introduced into the stream of neon at such a place that it did not have to pass through a charcoal tube before reaching the tube containing the wire. It was rapidly removed from the neon by the charcoal, and several admissions* were necessary before the accommodation coefficient reached the saturation value.

The results for two experiments at about 30° C are shown in fig. 1*a*, in which a, the accommodation coefficient of neon, is plotted against time, each admission of oxygen being shown by an arrow. The zero of time is the time at which the initial flashing current was cut off. The time scale is the same for both experiments, but each has its own scale of a. It should be mentioned that small changes in the accommodation coefficient under given conditions (current, gas pressure, roughness of surface) can be detected with certainty, even when they are considerably less than the uncertainty in the absolute value. It will be seen that in each experiment a appears to approach a limiting value. This indicates saturation of the oxygen adsorbed film.

---

* Each admission would have been sufficient in the absence of the charcoal to produce a pressure of about 5×10^{-4} mm of mercury in the whole apparatus of volume about three litres.

When this stage was reached the wire was heated for one minute by passing a current *i* through it, and after the current was cut off and the temperature had settled down again the accommodation coefficient was again measured. The results of this experiment are shown in fig. 1*b*, in which the accommodation coefficient is plotted as a function of the maximum temperature to which the filament was raised* by the current

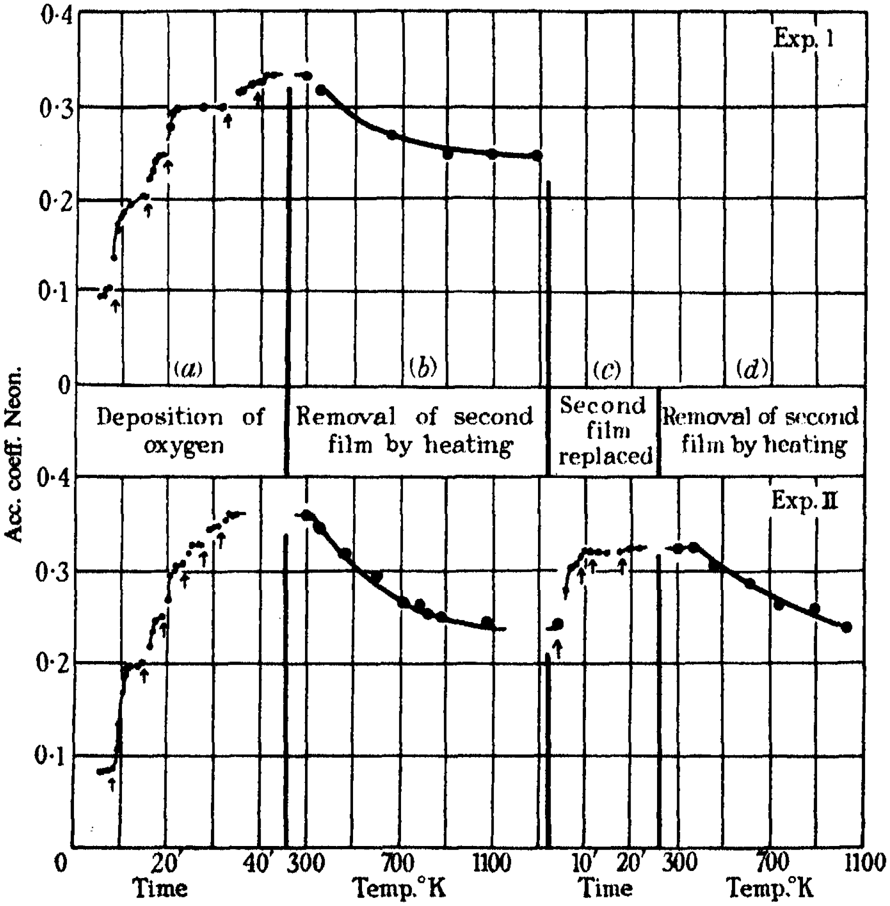


Fig. 1.

---

* This maximum (central) temperature was obtained as follows. The mean temperature of the filament was measured by measuring the resistance while the heating current *i* was passing. A rough estimate of the difference between the measured mean temperature and the central temperature was obtained, using the data given by Langmuir, MacLane, and Blodgett (' Phys. Rev.,' vol. 35, p. 478 (1930)) for tungsten filaments *in vacuo*. The difference varied from about 10° at 350° K to about 50° at 1050° K. At the low pressure of neon used (about 0·07 mm mercury) these figures certainly give the order of the correction.

i. The results in experiment II indicate an oxygen film which becomes unstable for times of the order of 1 minute at about 330° K. There are not enough low-temperature points in experiment I to fix this temperature, but it is included because there are more high-temperature points. These indicate another film which is stable up to 1300° K and probably higher. This latter film can be identified with the oxygen film which has been studied, using thermionic methods,* which is stable up to nearly 2000° K. The other film is new.† After it had been removed so that the accommodation coefficient had fallen to 0·24, it could be replaced by again admitting oxygen. This is shown in fig. 1*c*. In fig. 1*d* the removal of this film by heating again is shown, the temperature indicated for instability this time being 380° K. We take the mean of the two temperatures 355° K as the temperature at which the second film becomes unstable for times of the order of 1 minute.

2—HEAT OF ADSORPTION AND AMOUNT ADSORBED

The properties of both oxygen films have been studied, using the technique described for hydrogen. The numerical results obtained will be summarized in this section. There are certain technical difficulties in working with oxygen due to (*a*) diffusion of oxygen into the tungsten, (*b*) a slow take-up of oxygen probably on the glass at liquid air temperature or on mercury surfaces. These make the results less precise than those for hydrogen. Of these (*b*) was the more troublesome as (*a*) really only introduced slight difficulties of interpretation. The effect of (*b*) was that the deflection of the Pirani gauge was always less on evacuation than on admission and, further, the deflection on admission after the wire was saturated could not be obtained so definitely as with hydrogen. (*a*) in itself is very interesting and will be investigated in detail with apparatus designed for the purpose. It is discussed very briefly in considering the results given in Table I.

Before discussing Table I it will be convenient to give a short summary of quantities which have been measured and results obtained, as this will serve to indicate the scope of the experiments.

The amount of oxygen adsorbed in the first film shows that there is approximately one atom per atom of tungsten in the surface.

---

* *See* Langmuir and Villars, ' J. Amer. Chem. Soc.,' vol. 53, p. 486 (1931). It was removed by flashing in the usual way.

† It is quite different from an oxygen film noted by Blodgett and Langmuir (' Phys. Rev.,' vol. 40, p. 103 (1932)) which, in the presence of oxygen at a pressure of 10^{-8} mm, is unstable above 200° K.

The heat of adsorption for the first film has been determined, and is in reasonable agreement with that deduced from Langmuir and Villars's study of the evaporation of oxygen from tungsten.

TABLE I

Experiment	Admission	Notes	Molecules O_2 adsorbed in each admission	Heat of adsorption (K calories per mol O_2)
I	1	Complete adsorption	$1 \cdot 67 \times 10^{14}$	126
	2	,, ...	1·67	130
	3	Saturation reached	1·46	69
	4	Adsorption due diffusion between 3 and 4	1·03	36
	5	Adsorption due diffusion between 4 and 5	0·53	—
	6	No adsorption	0	—
	7	,, ...	0	—
	8	Wire heated to 1050° K to remove top film before 8	0·50	51
	9	No adsorption	0	—
II	1	Complete adsorption	$1 \cdot 87 \times 10^{14}$	134
	2	,, ...	1·87	112
	3	Saturation reached	1·32	72
	4	Adsorption due diffusion........	0·43	50
	5	,, ...	0·14	—
	6	No adsorption	0	—
	7	,, ...	0	—
	8	Wire heated to 1050° K	0·46	45
	9	No adsorption	0	—
III	1	Complete adsorption	$1 \cdot 33 \times 10^{14}$	145
	2	,, ...	1·33	155
	3	Saturation reached	1·21	100
	4	Adsorption due diffusion........	0·61	65
	5	,, ...	0·28	—
	6	,, ...	0·11	—
	7	No adsorption	0	—
	8	,, ...	0	—
	9	Wire heated to 1050° K	0·36	43
	10	No adsorption	0	—

The amount adsorbed in the second film has been determined with sufficient accuracy to make it possible to draw deductions concerning its nature.

The temperature at which this second film becomes unstable is known from the results given in § 1 and this, together with an approximate measurement of the heat of adsorption, shows that the oxygen is adsorbed as molecules.

As with hydrogen, the evidence as to the type or types of adsorption comes from both the heat of adsorption and the number of molecules adsorbed. Before analysing these two quantities in detail we shall consider the whole results in a general way and deduce from them and collect together the numerical values required for the subsequent discussion.

In the first place it will be seen that for the first two admissions in each experiment the heat of adsorption remains practically constant and high. Thus all or practically all these molecules must go into the first film. With the third admission in each case saturation is reached and the heat of adsorption is considerably lower. Now, when the residual oxygen is pumped out and another, lot 4, is admitted, further adsorption takes place with a lower heat of adsorption; therefore, presumably, the adsorption is in the second layer. Similarly adsorption occurs at admission 5 and in the case of experiment III at admission 6, the criterion of adsorption being the occurrence of an appreciable if not accurately measurable heating effect. After this no further adsorption takes place. This shows that the adsorption observed in 4 and 5 is not due to evaporation during evacuation of some of the molecules in the second film; for, if it were, the effect would persist indefinitely. It must therefore be due to the production of spaces in the second layer by some process of diffusion of oxygen, the place to which it diffuses becoming saturated. The nature of this diffusion will be considered later.

Admission 4 gives a measure of the heat of adsorption in the second film. Using this and the heat of adsorption in the first film given by admission 2, it is possible to estimate in each experiment from the measured heat in 3 what proportion of the molecules adsorbed in 3 goes to the first and what to the second film. Thus the total numbers of molecules adsorbed in the first and second films up to saturation (*i.e.*, after admission 3) can be estimated and are given* in columns 2 and 3 of Table II, the ratio of the numbers being given in the fourth column.

Another measure of the heat of adsorption and number of molecules in the second film is given by the admission 8 or 9 after the second film had been removed by heating the wire. The heats obtained in this way, 51, 45, and 43 kilocalories, are in reasonable agreement with those of 36,

---

* It should be mentioned that the adsorption in 4 of experiment I is greater than the number of molecules in the second film, but it may include some diffusion.

50, and 65 given by admission 4. The mean of them all is 48 kilocalories per mol O_2 with an R.M.S. deviation of ± 10. It should be mentioned that the uncertainty in the total amount of gas admitted mentioned before affects the accuracy of these six determinations, where the amount of residual gas is considerable, much more seriously than it does that of the first three admissions where all or nearly all the gas admitted is adsorbed. The ratio of the number of molecules adsorbed in admissions 8 or 9 from Table I to the number in the first layer is given in the last column of Table II.

TABLE II

Experiment	Number of molecules		Ratio second to first layer	
	First layer	Second layer	From 3	From 8 or 9
I	$3 \cdot 85 \times 10^{14}$	$0 \cdot 95 \times 10^{14}$	0·25	0·13
II	4·22	0·85	0·20	0·11
III	3·14	0·74	0·24	0·11

3—THE FIRST OXYGEN FILM

The number of oxygen molecules adsorbed in the first film (*see* Table II) is almost the same as the number of hydrogen molecules adsorbed on the same surface (p. 453), so that there is one atom of oxygen for each atom of tungsten in the surface. It has always been assumed that this stable oxygen film is an atomic film, and there is a fairly strong indication in the phenomenon described later in another paper* that this is so.

The direct measurement of the heat of adsorption is consistent with this view. Taking the mean of the six values (126, 130, 134, 112, 145, 155) given in Table I for the first two admissions, we obtain 134 kilocalories per mol O_2. The numerical test described in § 5 showed that in the first two admissions about 5% of the total molecules adsorbed go into the second film. Taking 48 kilocalaries per mol O_2 as the heat of adsorption in this film, we have, if y is the heat of adsorption in the first film,

$$0 \cdot 95y + 0 \cdot 05 \times 48 = 134.$$

Thus the heat of adsorption in the first film is 139 kilocalaries per mol O_2. For reasons already given, these experiments are not sufficiently accurate to determine the variation of heat of adsorption as the surface becomes progressively more covered, and this is the mean value from $\theta = 0$ to $\theta = 1$.

---

* This volume, p. 478.

The heat of dissociation of oxygen is 117 kilocalories* per mol O_2. The mean heat required to remove an atom from the surface would therefore, according to the measurement of the heat of adsorption, be 128 kilocalories per gram-atom. Langmuir and Villars, from a study of the rate of evaporation, obtain 162 kilocalories per gram atom for the heat of desorption. This refers to a surface which is very sparsely covered, and Langmuir states that the rate of evaporation shows that the heat of desorption decreases progressively and considerably as the surface becomes covered. Thus the mean value would be considerably below 165 and would probably be fairly close to 128, the present directly determined value.

It must be mentioned that in this particular case the heat measurements would be equally consistent with the view that the oxygen is adsorbed and evaporates as molecules.

4—STABILITY AND HEAT OF ADSORPTION OF THE SECOND OXYGEN FILM

In § 2 it has been shown that the measured heat of adsorption of the second film is 48 kilocalories per mol O_2 with a R.M.S. deviation of $\pm$ 10. From the evaporation formula

$$t_f = -2\cdot06\ .\ 10^{-13}(\log_{10} f)(M^{\frac{1}{2}}T/^{\frac{1}{2}}\Phi)\,e^{\Phi/RT},$$

already used for hydrogen atoms,† values of $\log_{10} t_{\frac{1}{2}}$ for oxygen molecules ($M = 32$) have been calculated at various temperatures for $\Phi = 30$, 40, and 50 kilocalories per mol and are plotted against $1/T$ in fig. 2.

The results given in fig. 1 show that the second oxygen film becomes unstable for times of the order of 1 minute at about 360° K. Fig. 2 shows that for a heat of desorption just under 30 kilocalories per mol $\log t_{\frac{1}{2}}$ is of the order 2 in this temperature range. Thus, according to the evaporation formula, the heat of desorption of the second film is around 30 kilocalories per mol. This is in reasonable agreement with the directly measured value of 48, as an uncertainty of 30% in the measured heat would account for the difference, and an error of this order in the absolute value is possible.

If, on the other hand, the film were atomic, the heat of desorption would certainly be above 60 kilocalories per gram atom and, using the measured heat of adsorption, would be 82 kilocalories. Such a film would be stable up to much higher temperatures. We conclude, therefore, that the oxygen in the second film is adsorbed as molecules. The high heat of adsorption indicates that the binding to the surface is chemical in type.

---

* Hinshelwood and Williamson, "Reaction between Hydrogen and Oxygen," p. 26 (Oxford, 1934).

† *See* p. 460.

5—THE NUMBER OF MOLECULES IN AND NATURE OF THE SECOND FILM. STRUCTURE OF THE FIRST FILM

We have seen that the number of oxygen atoms in the first film would correspond to an atomic film of the same spacing as that of tungsten. When account is taken of the second film being molecular, the amount of oxygen in it is surprisingly small for the following reasons. The distance between the atoms in the oxygen molecule is 2×10^{-8} cm, which is less than the distance between the oxygen atoms in the first film. If, then, a second film is formed at all, general considerations would make it seem very likely that the spacing in it would bear some simple relation to that in the first film and very unlikely that the ratio of the two spacings would be greater than two. That is to say, in the second film there would be one-quarter or more the number of particles or, since it is molecular and the first film atomic, one-half the number of atoms. The number found is less than one-quarter (*see* Table II, column 4) and, as is pointed out in a footnote later in this section, this is probably a slight over-estimate. The explanation of this is, in fact, very simple and is an illustration of an important and inherently necessary property of any film which, like the first oxygen film, is formed by the adsorption on neighbouring solid atoms of the two atoms of diatomic gas molecules and in which the adsorbed atoms are immobile and stable. Such a film is *necessarily* imperfect and incomplete, having gaps or holes in it.

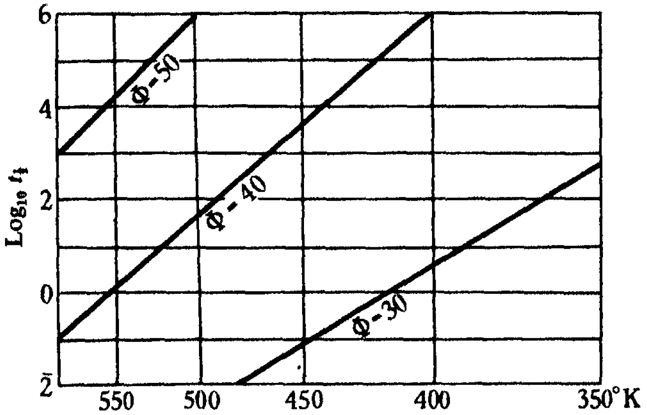


FIG. 2.

For the sake of definiteness, let us consider adsorption on a plane in which each surface atom is surrounded by four equally spaced other atoms. As the film is gradually built up, certain single surface atoms will find themselves surrounded by four filled places. Such atoms will be able to take no part in the adsorption process and will remain bare. The complete film thus of necessity has a sort of irregular mosaic structure.

To find what proportion of the surface atoms remain bare an experimental test has been carried out in which neighbouring pairs of points on a diagram of a (110) plane of tungsten have been selected at random* and occupied. The result is shown in fig. 3, in which the occupied places are shown by small circles and the spaces that remain bare are shown

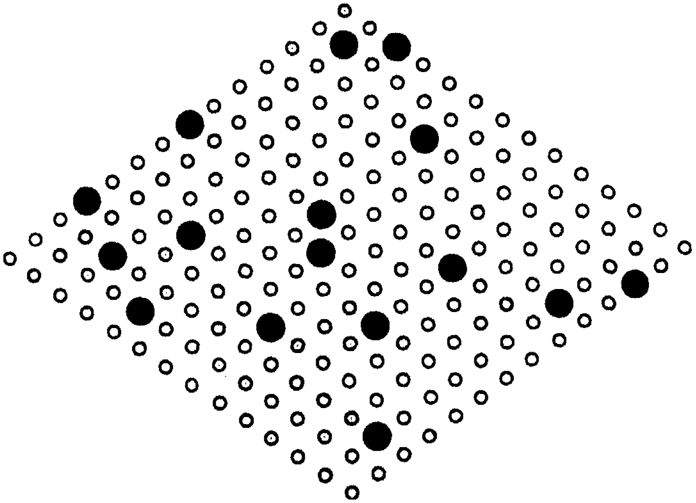


FIG. 3.

---

* The method used was as follows. Each point may be represented by two numbers (oblique co-ordinates). Cards numbered 1 to 20 were used. The position of one atom in the impinging molecule was fixed by picking a card, replacing it and shuffling, and then picking another. The position of the second atom was found by choosing in a similar way a number (1 to 4), giving its orientation relative to the first. If both spaces were occupied the molecule was taken to be "reflected," but if either space was unoccupied the atom striking it was assumed to stick and the other to rotate about it in a clockwise direction till it found an empty neighbouring space. In view of what will be said later about the adsorption of molecular oxygen, this seemed to be the most satisfactory procedure. In this way a block of 400 places was filled up and a number of places in the next peripheral line outside these. In order to exclude disturbing effects due to boundary only the middle 196 places have been included in fig. 3, that is three complete peripheral lines have been neglected. The inclusion of two of these would not have affected the numerical result.

dark. Out of the total 196 places 16 are unoccupied, that is 8·2% of the whole, while of the 100 central places 8 are unoccupied. We may therefore take it that about 8% of the total number of surface atoms remain bare.

In order to test whether the particular assumption about the atomic arrangement is of importance, a similar test was carried out on a hexagonal arrangement, so that there could be six possible places for the second atom of a molecule to occupy. The numerical result was identical, 8 gaps in the central 100 places and 16 gaps in the central 196 places.

These uncovered tungsten atoms will undoubtedly exert a much greater attraction on impinging oxygen molecules than the other parts of the surface, and it seems certain that the second film must consist of oxygen molecules adsorbed above them. In other words, for the adsorption of oxygen molecules these will be active parts of the surface.* Thus for 92 atoms in the first film we should expect to find 8 molecules or 16 atoms in the second film, that is the ratio of the number of atoms in the second and first films would be 0·17. The difference between this and 0·23, the mean of the experimental values given in column 4 of Table II, is only 25% of the latter.† The agreement must be regarded as very significant experimental evidence as to the correctness of the present view concerning the structure of the first oxygen film. This view arises in an entirely natural and unforced way when the process of building up the film is considered in detail.

Other types of film will have similar gaps. For example, an immobile film formed without dissociation from a gas if the spacing of the adsorbed particles is double that of the underlying solid atoms.

6—ACCOMMODATION COEFFICIENT OF OXYGEN-COVERED SURFACE

We must now consider in the light of these results the numerical values for the accommodation coefficients of the oxygen-covered surface. Fig. 1 shows that the removal of the second film changes the accommodation coefficient of neon from about 0·36 to 0·24. With a bare surface it is 0·08. Let a_1 be the accommodation coefficient of neon atoms which strike adsorbed oxygen atoms and a_2 that of those which strike the adsorbed oxygen molecules. Assuming that the resultant accommodation coefficient is obtained by simple proportion, we have, since in the first film 8% of the total number of places are gaps,

---

* They would probably also in some cases be centres of catalytic activity.

† It would, in fact, be expected that the numbers in column 4 of Table II would be slightly greater than the number of molecules in the second film, if the diffusion already mentioned took place to an appreciable extent during the few seconds taken to make the thermal measurements after the admission of the oxygen.

$$0\cdot08 \times 0\cdot08 + 0\cdot92 \times a_1 = 0\cdot24,$$

so that

$$a_1 = 0\cdot25;$$

$$0\cdot08 \times a_2 + 0\cdot92 \times 0\cdot25 = 0\cdot36,$$

so that

$$a_2 = 1\cdot5,$$

which is impossible, since the accommodation coefficient cannot be greater than unity.

In making this calculation we have effectively assumed that, if in fig. 4 the adsorbed molecule is represented by the two crosses and the surrounding adsorbed atoms by the circles, the effect of the molecule only covers the shaded region. Actually this is not so, and all neon atoms colliding with the surface inside the hexagon will be affected more or less by the presence of the molecule. The total area of the hexagon is three times that of the shaded portion. We shall assume that for neon atoms striking the shaded portion the accommodation coefficient is a_2, and as a rough approximation that for those striking the rest of the hexagon it is on the average $(a_2 + a_1)/2$. Similar considerations will apply when the place occupied by the molecule is bare. In this latter case we have, remembering that the accommodation coefficient of a bare surface is 0·08,

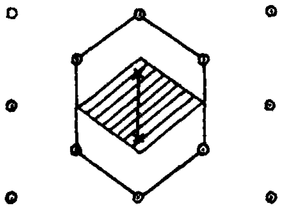


FIG. 4.

$$0\cdot08 \times 0\cdot08 + 0\cdot16\,(a_1 + 0\cdot08)/2 + 0\cdot76\,a_1 = 0\cdot24,$$

so that

$$a_1 = 0\cdot27.$$

When the spaces are occupied by molecules,

$$0\cdot08 \times a_2 + 0\cdot16\,(a_2 + 0\cdot27)/2 + 0\cdot76 \times 0\cdot27 = 0\cdot36,$$

from which

$$a_2 = 0\cdot7.$$

It may be supposed that the increased power of an adsorbed molecule over an atom in communicating energy to or taking it from an impinging gas atom is due in part to its being able to rotate.

The theory of energy interchange between gas atoms and solids* shows that the value of the accommodation coefficient depends on the balance between a number of factors, the masses of the gas and solid atoms, the characteristic frequency of the solid, the law of force, both attractive and repulsive, between the gas and solid atoms. It is unwise to stress too much, as is sometimes done, the importance of any one of these; for example, the existence of unsaturated surface forces. In the present case the surface covered with oxygen has a much higher accommodation coefficient than the bare unsaturated surface of tungsten.

7—NOTES ON THE DIFFUSION OF OXYGEN INTO TUNGSTEN

The detailed nature of the diffusion process which causes the adsorption shown in Table I subsequent to initial saturation will be elucidated only when the new apparatus already mentioned has made it possible to study it fully. Certain deductions can, however, be made from the results given.

The first question is whether the diffusion takes place from the first or the second film. If it took place from the first film, atoms from this would pass into the solid and leave single gaps where, on subsequent admission of oxygen, molecules would be adsorbed. The total number of molecules in the second film would be considerably increased by this process. All these molecules would be removed by heating the wire before admissions 8 or 9 in Table I, and it would be expected that the adsorption of gas in these admissions would be considerably in excess of the initial number of molecules in the second film. The figures in the last two columns of Table II show that this is not what happens and that, in fact, the numbers in the last column are appreciably less than those in the fourth column.† It must therefore be supposed that the diffusion takes place from the second film. The gaps in the first film thus have a further significance in that they must be regarded as the points at which the attack on the metal lattice by oxygen first takes place. The actual number of molecules which penetrate at this first stage is not yet known, but the process is a comparatively rapid one and then saturation is reached, or, what is more likely, the further process of penetration is much slower so that saturation appears to be reached. There is no diffusion after admissions 8 or 9.

---

* For example, Jackson and Mott, 'Proc. Roy. Soc.,' A, vol. 137, p. 703 (1932).

† It has already been mentioned that the numbers in the fourth column are probably an over-estimate. The numbers in the last column should strictly be compared not with them but with 0·17, the theoretical ratio deduced in § 5. The difference would be covered by an error of 30% in the experimental values.

This suggests that the atoms which have diffused are not seriously affected by heating to 1000° K, so that they are tightly bound to the lattice and do not evaporate and also the further process of diffusion remains slow at this temperature. It will be interesting with the new apparatus to find the temperature at which they are affected. The stability would appear to suggest that the oxygen is bound in the atomic form, presumably to the layer or layers of tungsten atoms below the surface layer.*

SUMMARY

The methods described in the foregoing paper have been used to study the adsorption of oxygen on tungsten. The heat of adsorption and amount of oxygen in the well-known stable film have been measured. A second molecular film has been found and studied in a similar way.

The process of building up an immobile film by the adsorption of the two atoms of a diatomic molecule on neighbouring solid atoms is considered in detail and it is shown that such a film necessarily has gaps in

---

* This reasoning could, but would not necessarily, be invalidated if the first film became mobile at 1050° K, the temperature at which the second film is removed (Becker, 'Phys. Rev.,' vol. 34, p. 1333 (1929), has shown that it is mobile at 1400° K). Let us consider the experiment in detail. The wire is raised to 1050° K for 1 minute and the molecules are removed, leaving gaps in the atomic film. If this film is mobile at this temperature, the gaps will of course move about, and at times two gaps will come to occupy neighbouring places. If this occurs an appreciable number of times during the first few seconds of heating (*i.e.*, before the gas evaporated has been completely removed by the pumps), these gaps will have an opportunity of becoming filled; so that when, after the wire has cooled down, more gas is admitted (admissions 8 or 9) the amount adsorbed into the second film will be less than the number of molecules in this film before evaporation. If, as has been postulated in this section, diffusion is from the second film, a simple kinetic calculation shows that this mobility would not affect appreciably the adsorption in admissions 8 or 9—account must be taken of the fact that the gaps in the first film repel each other. Thus the agreement between the observed adsorption in 8 or 9 and the initial of gaps in the atomic film would still be favourable to the view that all the diffusion takes place from the second film. On the other hand, if the diffusion takes place from the atomic film, we have already noticed that the number of gaps produced when the second film evaporates would be considerably greater than the initial number. In this case the kinetic calculation becomes complicated; but it is obvious from what has been said that the mobility of the first film would leave a possible loophole for explaining the small adsorption observed in admissions 8 or 9. The agreement between the amount of this adsorption and the initial number of gaps would have to be regarded as a coincidence. No doubt the detailed quantitative study will make a definite decision possible. The reservation contained in this footnote does not, of course, affect in any way what has been said in § 5.

it. The number of these gaps has been obtained by an experimental numerical test, and it has been shown that the amount of oxygen in the second molecular film would correspond to adsorption on the gaps in the first film. Other properties of these gaps are discussed, including their possible role in the diffusion of oxygen into tungsten, which has been observed in the experiments.

The accommodation coefficients of neon striking the atomic and the molecular film are deduced.

Composite Films of Oxygen and Hydrogen on Tungsten

By J. K. Roberts, Laboratory of Colloid Science, Cambridge

(*Communicated by E. K. Rideal, F.R.S.—Received June 5—Revised July 17, 1935*)

The methods described in the two foregoing papers have been applied to the study of the adsorption of oxygen on a hydrogen-covered surface and of the behaviour of hydrogen when the surface has previously been covered, or nearly covered, with oxygen.

1—THE ADSORPTION OF OXYGEN ON A HYDROGEN-COVERED SURFACE

Using the accommodation coefficient of neon as an indicator, it has been shown that oxygen can be adsorbed at room temperature on a tungsten surface already covered with hydrogen. The bare tungsten surface was first covered with hydrogen so that the accommodation coefficient rose to 0·16. Oxygen was then introduced and after several admissions the accommodation coefficient rose to 0·36.

This adsorption was investigated, using the apparatus for measuring heats of adsorption. The bare wire was completely covered with hydrogen. A charge of oxygen of $1 \cdot 2 \times 10^{14}$ moles was then admitted (*cf.* Table I of the previous paper). The deflection of the Paschen galvanometer showed that an amount of heat was developed in the wire of the same order as in admission 1, experiments I, II, and III in this table. At the same time the Pirani gauge showed a large deflection, in some cases slightly larger than would correspond to the total amount of oxygen in the charge as determined later after all adsorption was completed. We thus have a large heat of adsorption of the same order as would correspond to the adsorption of the complete charge on a bare tungsten surface, while at the same time the pressure measurements indicate no adsorption at all.

The only possible explanation of this is that the oxygen is adsorbed on the tungsten and throws the hydrogen off into the gas phase. The gas cannot be water vapour, as this would be taken up by the liquid air traps, the vapour pressure of ice at liquid air temperatures being quite negligible. The pressure developed when the oxygen was admitted corresponded to $0 \cdot 9 \times 10^{14}$ moles, assuming the gas to be hydrogen, and the total charge of oxygen was $1 \cdot 2 \times 10^{14}$ moles, so that roughly speaking each molecule of oxygen adsorbed throws off a molecule of hydrogen.* Since the hydrogen is adsorbed as atoms, it is difficult to visualize this process unless the oxygen is also adsorbed as atoms.

Consider the energy balance in this process. The measured heat of adsorption for the process is of the order 100 to 130 kilocalories per mol O_2. The heat of desorption of hydrogen molecules (*i.e.*, the measured heat of adsorption) is about 20 kilocalories per mol H_2 for a covered surface. The total amount of heat, 120 to 150 kilocalories per mol, is of the order that can be made available by the adsorption of oxygen on tungsten. If, on the other hand, the hydrogen were thrown off as two separate atoms, the heat of desorption† would be 120 kilocalories per mol H_2. Adding this to the measured heat of adsorption for the process, we obtain altogether 220 to 250 kilocalories per mol. This amount of heat cannot be made available by the adsorption of oxygen on tungsten. The hydrogen could only be thrown off as atoms if the heat of adsorption of an oxygen molecule on two neighbouring bare spaces completely surrounded by places occupied by hydrogen atoms were much greater than that of oxygen on a bare surface. If this were so, the heat of adsorption of hydrogen on a surface partially covered with oxygen would be greater than that of hydrogen on a bare surface. This has been measured and has been shown, in fact, to be less than that on bare tungsten (the measured values were 20 to 30 kilocalories per mol H_2). We conclude, therefore, that the hydrogen must be thrown off as molecules and not as separate atoms, and also that the presence of hydrogen atoms in neighbouring places on the surface diminishes the heat of adsorption of oxygen on bare tungsten atoms. It can further be concluded from this measurement of the heat of adsorption of hydrogen on a surface nearly covered with oxygen that the dipoles for hydrogen- and oxygen-adsorbed atoms are of the same sign, *i.e.*, the binding of hydrogen to the tungsten surface is not similar to that of adsorbed films of the alkali metals. Measurement of the electrical properties of the hydrogen-adsorbed film will be interesting in this connexion.

---

* It may be mentioned that the gas behaved like hydrogen and did not disappear slowly as oxygen always did.

† *See* p. 459.

In order to account for certain facts in connexion with the thermionic emission from composite surfaces Becker* has suggested that when the temperature is raised a tungsten-barium-oxygen surface stable at room temperature switches over into a tungsten-oxygen-barium surface (similarly with caesium substituted for barium); but he pointed out that there is no direct evidence for this view. The present results may be held in a sense to provide by analogy such direct experimental evidence, since they show that oxygen can displace another substance which is adsorbed on tungsten. Whether the displaced substance remains adsorbed on top of the oxygen or whether it is thrown right off, as in the present work, is a matter of detail.

Up to the present we have considered only the first admission of oxygen. The results obtained for the later admissions are difficult to interpret with certainty as the proportion of hydrogen to oxygen in the residual gas is not known. The analysis of this gas will be possible by combining an absolute manometer with the Pirani gauge. This apparatus is at present being made, and it is hoped that with it very soon definite evidence will be available as to the amount of oxygen that can be adsorbed and the amount of hydrogen displaced; that is, it will be possible to see, for example, whether the state finally reached is one in which the surface is completely or almost completely covered with oxygen, or one in which the amounts of the two substances in the first film are of the same order. In either case the consideration of the formation of the film from the point of view of § 5 of the previous paper will give significant results. These will be discussed elsewhere.

At room temperature the displacement of hydrogen takes place extremely rapidly, and it is not possible to say whether or not the oxygen is first adsorbed on top of the hydrogen before the displacement occurs. The experiments with the new apparatus will be carried out at lower temperatures to see if a condition analogous to the tungsten-barium-oxygen surface can be found in which the oxygen is adsorbed and remains on top of the hydrogen and then, if the temperature is raised, displaces it.

---

* Becker, 'Phys. Rev.,' vol. 34, p. 1325 (1929).

2—THE BEHAVIOUR OF HYDROGEN WITH A SURFACE PARTIALLY COVERED WITH OXYGEN

It has been mentioned in the preceding section that the heat of adsorption of hydrogen on a surface partially covered with oxygen has been measured. In carrying out these experiments an effect was observed which may be of importance from the point of view of the problem of activated adsorption. In several experiments, when the hydrogen was admitted after partially covering the surface with oxygen, the Pirani gauge showed that the whole amount was not adsorbed but that there was a small but appreciable amount of residual gas. The gauge showed that this gas gradually disappeared and this disappearance was fully confirmed by opening the system to the pumps after a few minutes, when no deflection of the gauge occurred, showing that the disappearance was complete.

A detailed quantitative study of this disappearance will be made with the new apparatus, but it may be said that, whatever the explanation may be, it can hardly fail to be of interest. It has already been mentioned in § 7 of the previous paper that diffusion of oxygen at room temperature from the first adsorbed layer cannot be completely excluded as a possibility. If this occurs, neighbouring gaps could be left on which hydrogen from the gas phase would be adsorbed. If this should be the process involved the quantitative study of the diffusion of oxygen would be considerably facilitated.

On the other hand, it is a striking thing that by covering or partially covering the tungsten surface with oxygen we have produced conditions under which phenomena are observed which are similar to the phenomenon of activated adsorption. Now it has already been shown* that the tungsten surfaces on which activated adsorption has been observed cannot initially have been bare tungsten. It is known, too, that activated adsorption occurs on various oxide surfaces. Taking these three points together, it seems legitimate to suggest tentatively that activated adsorption of hydrogen is connected with the interaction between hydrogen and surface oxygen, either adsorbed or forming an integral part of the solid lattice.

SUMMARY

The adsorption of oxygen on a surface already covered with hydrogen has been studied and it has been shown that when an oxygen molecule is adsorbed a hydrogen molecule is thrown off. The behaviour of hydrogen in the presence of a surface partially covered with oxygen is also considered. The bearing of the results on the problem of activated adsorption is discussed.

---

* *See* p. 462.

On the Radiative Collision between Fast Charged Particles

By C. Møller†

(*Communicated by R. H. Fowler, F.R.S.—Received June* 5, 1935)

Introduction

The radiation emitted in the collision of a fast charged particle with a nucleus has been treated in quantum mechanics by Heitler, Bethe, and various other authors.‡ In all cases the nucleus has been considered infinitely heavy, the field of the nucleus being then an external field so that the problem reduces to a one-body problem. It may be of some interest to consider the more general case, where the two colliding particles have masses comparable with each other. When a fast particle traverses matter many different elementary processes are possible. Some of these are: (1) collisions with the atomic electrons; (2) collisions with the atomic nucleus in both cases without emission of radiation; we further have (3) radiative collisions with the atomic electrons; and (4) radiative collisions with the nucleus. If the colliding particle is an electron the third process is of the type we wish to consider where the masses of the particles are comparable.

For not too large velocities of the incident particle the stopping of the particle is mainly due to the first process, which has been treated in quantum mechanics by various authors.§ For very large velocities and large atomic number the fourth process is the most important.‖ If, however, the velocity is very large and the atomic number small the third process will presumably give a contribution to the stopping which is large compared with that of the first and comparable with that of the fourth process. Further, when one has treated the radiative collision of two electrons, one can immediately answer two other questions. If one of the electrons after the collision is in a negative state the collision process is the converse of the photo-production of a pair of positive and negative electrons in the presence of an electron. And if both electrons are in a negative state after the collision we have the converse process of the simultaneous production of two pairs by a photon. (The last process is only possible if the particles are not free.)

---

† Rockefeller Fellow temporarily in Cambridge.

‡ Heitler, 'Z. Physik,' vol. 84, p. 145 (1933); Bethe and Heitler, 'Proc. Roy. Soc.,' A, vol. 146, p. 83 (1934); v. Weizsäcker, 'Z. Physik,' vol. 88, p. 612 (1934); Williams, 'Phys. Rev.,' vol. p. 45, 729 (1934).

§ Bethe, 'Z. Physik,' vol. 56, p. 293 (1932); Williams, 'Proc. Roy. Soc.,' A, vol. 135, p. 108 (1932); Møller, 'Ann. Physik,' vol. 14, p. 531 (1932), later referred to as II.

‖ Bloch, 'Z. Physik,' vol. 81, p. 363 (1933); 'Ann. Physik,' vol. 16, p. 285 (1933).

The problem of the radiative collision of two particles can be treated with the quantum electrodynamics of Heisenberg and Pauli† and Dirac.‡ These theories, however, involve the well-known difficulties of the infinite self-energy of the particles, and to get physical results from the theory one has afterwards to subtract infinite terms. We therefore propose another method which allows us in an elementary way to treat the radiative collision independently of quantum electrodynamics. The adopted method is a correspondence method which forms an immediate generalization of the method previously used in the treatment of the non-radiative collision of fast particles.§

1—We distinguish the two colliding particles by the suffixes 1 and 2.

Let us suppose that both particles separately obey a Dirac equation having the masses m_1, m_2, and the charges $-e_1$ and $-e_2$ respectively. For free particles the eigenfunctions are of the form

$$\left.\begin{aligned} \psi_{n_1} &= \phi_{n_1} e^{-i\mathrm{E}_1 t/\hbar} = a_1 e^{i\left[\frac{1}{c}(\mathbf{p}_1 . \mathbf{x}_1) - \mathrm{E}_1 t\right]/\hbar} \\ \psi_{n_2} &= \phi_{n_2} e^{-i\mathrm{E}_2 t/\hbar} = a_2 e^{i\left[\frac{1}{c}(\mathbf{p}_2 . \mathbf{x}_2) - \mathrm{E}_2 t\right]/\hbar} \end{aligned}\right\} . \tag{1}$$

Here the label n_1 indicates a state in which the first particle has the momentum $\mathbf{p}_1/c$ the energy E_1 and a definite direction of spin. There are four eigenfunctions with the same value of $\mathbf{p}_1$ corresponding to the two different directions of spin and the double sign of the energy. The amplitude a_1 has four components and is supposed to be suitably normalized. It is practical to imagine the particles enclosed in a box of finite volume V and afterwards to let V become infinite. In this case $\mathbf{p}_1$ takes on only discrete values and we can put

$$a_1 = \frac{1}{\sqrt{\mathrm{V}}} u_1 \qquad \text{with} \qquad \bar{u}_1 u_1 = 1. \tag{2}$$

The same holds for the second particle.

---

† 'Z. Physik,' vol. 56, p. 1 (1929); vol. 59, p. 168 (1929).

‡ 'Proc. Roy. Soc.,' A, vol. 136, p. 453 (1932).

§ Møller, 'Z. Physik,' vol. 70, p. 786 (1931). Later referred to as I.

Let $n_1^0 = (\mathbf{p}_1^0, \mathrm{E}_1^0)$ and $n_2^0 = (p_2^0, \mathrm{E}_2^0)$ be the initial states of the two particles. We ask for the probability of a transition in which particle 1 goes over into the state $n_1 = (\mathbf{p}_1, \mathrm{E}_1)$ and particle 2 goes over into $n_2 = (\mathbf{p}_2, \mathrm{E}_2)$, while a photon of frequency $\nu = (\mathrm{E}_1^0 + \mathrm{E}_2^0 - \mathrm{E}_1 - \mathrm{E}_2)/\hbar$ is emitted into the solid angle $d\Omega_k$ in the direction of the unit vector $\mathbf{n}$. In a transition $(n_1^0, n_2^0) \rightarrow (n_1\, n_2)$ without emission of radiation treated in I one forms the charge and current density corresponding to the transition $n_1^0 \rightarrow n_1$ of the first particle. The corresponding electromagnetic field is then treated as a perturbation of the second particle which causes the transition $n_2^0 \rightarrow n_2$. The calculation is formally analogous to a photo-effect.

It is now a natural generalization of the method used in I to treat the present case, where a photon is emitted during the collision, analogously to the case of light scattering which can be treated by the correspondence method of Klein.† Accordingly we have to consider the change in the current distribution of the one particle caused by the perturbing field originating from the transition of the other particle and to calculate the electromagnetic radiation corresponding to this change in the current density. We shall now state the theory in detail and later discuss the question of the uniqueness of the results obtained.

The real charge and current density corresponding to a transition between the states n_1° and n_1 is given by

$$\rho_{n_1 n_1^0} + \rho_{n_1^0 n_1} = \rho_{n_1 n_1^0} + \bar{\rho}_{n\, n_1^0}$$
$$\mathbf{J}_{n_1 n_1^0} + \mathbf{J}_{n_1^0 n_1} = \mathbf{J}_{n_1 n_1^0} + \bar{\mathbf{J}}_{n_1 n_1^0}$$

with

$$\left.\begin{aligned} \rho_{n_1 n_1^0} &= -\, e_1 \bar{\phi}_{n_1} \phi_{n_1^0} e^{-i(\mathrm{E}_1^0 - \mathrm{E}_1)t/\hbar} \\ \mathbf{J}_{n_1 n_1^0} &= c e_1 \bar{\phi}_{n_1} \boldsymbol{\alpha} \phi_{n_1^0} e^{-i(\mathrm{E}_1^0 - \mathrm{E}_1)t/\hbar} \end{aligned}\right\} . \qquad (3)$$

The corresponding scalar and vector potentials are determined by the equations

$$\left.\begin{aligned} \Delta\Phi - \frac{1}{c^2}\frac{\partial^2\Phi}{\partial t^2} &= -\,4\pi\rho \\ \Delta\mathbf{A} - \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} &= -\,\frac{4\pi\mathbf{J}}{c} \\ \operatorname{div}\mathbf{A} + \frac{1}{c}\frac{\partial\Phi}{\partial t} &= 0 \end{aligned}\right\} , \qquad (4)$$

where the right-hand sides are given by (3).

---

† 'Z. Physik,' vol. 41, p. 407 (1927).

For free particles we get the following expressions for the potentials (*cf.* I):

$$\left.\begin{aligned} &\Phi_1 e^{2\pi i\nu_1 t} + \bar{\Phi}_1 e^{-2\pi i\nu_1 t} \\ &\mathbf{A}_1 e^{2\pi i\nu_1 t} + \bar{\mathbf{A}}_1 e^{-2\pi i\nu_1 t} \end{aligned}\right\}, \tag{5}$$

with

$$h\nu_1 = \mathrm{E}_1 - \mathrm{E}_1{}^0, \tag{6}$$

$$\left.\begin{aligned} \Phi_1 &= -\frac{h^2c^2e_1}{\pi}\frac{\bar{a}_1 a_1 e^{i(\mathbf{p}_1{}^0 - \mathbf{p}_1, \mathbf{x})/\hbar c}}{|\mathbf{p}_1{}^0 - \mathbf{p}_1|^2 - (\mathrm{E}_1{}^0 - \mathrm{E}_1)^2} \\ \mathbf{A}_1 &= \frac{h^2c^2e_1}{\pi}\frac{\bar{a}_1 \boldsymbol{\alpha} a_1{}^0 e^{i(\mathbf{p}_1{}^0 - \mathbf{p}_1, \mathbf{x})/\hbar c}}{|p_1{}^0 - p_1|^2 - (\mathrm{E}_1{}^0 - \mathrm{E}_1)^2} \end{aligned}\right\}. \tag{7}$$

If one of the colliding particles are not free but subject to an external field the eigenfunctions of this particle are not plane waves and we have to take instead of (7)

$$\left.\begin{aligned} \Phi_1 &= -e_1\int\frac{\bar{\phi}_{n_1}(\mathbf{x}_1)\,\phi_{n_1{}^0}(\mathbf{x}_1)\,e^{i(\mathrm{E}_1{}^0 - \mathrm{E}_1)|\mathbf{x}-\mathbf{x}_1|/\hbar c}}{|\mathbf{x}-\mathbf{x}_1|}\,d\mathbf{x}_1 \\ \mathbf{A}_1 &= e_1\int\frac{\bar{\phi}_{n_1}(\mathbf{x}_1)\,\boldsymbol{\alpha}\,\phi_{n_1{}^0}(\mathbf{x}_1)\,e^{i(\mathrm{E}_1{}^0 - \mathrm{E}_1)|\mathbf{x}-\mathbf{x}_1|/\hbar c}}{|x - x_1|}\,d\mathbf{x}_1 \end{aligned}\right\}. \tag{8}$$

If we introduce the expressions (1) for the functions ϕ_n in (8) it becomes identical with (7) (*cf.* II, p. 543, 544). It is easily seen that (8) is a solution of (4) in the more general case.†

We now have to consider the influence of the potentials (5) on the particle 2. Temporarily dropping the suffix 2 we write the Dirac equation for this particle in the form

$$i\hbar\frac{\partial\psi}{\partial t} = \mathrm{H}\psi - \mathrm{U}_1 e^{2\pi i\nu_1 t}\,\psi - \mathrm{U}^*{}_1 e^{-2\pi i\nu_1 t}\,\psi, \tag{9}$$

where

$$\left.\begin{aligned} \mathrm{U}_1 &= e_2\,\Phi_1 + e_2(\boldsymbol{\alpha}, \mathbf{A}_1) \\ \mathrm{U}^*{}_1 &= e_2\bar{\Phi}_1 + e_2(\boldsymbol{\alpha}, \bar{\mathbf{A}}_1) \end{aligned}\right\}, \tag{10}$$

and H is the hamiltonian without interaction.

Following the method given by Klein (*loc. cit.*) for the case of an external field we try to find approximate solutions of (9) of the form

$$\chi_m = \phi_m e^{-i\mathrm{E}_m t/\hbar} + \mathrm{F}_m e^{-i(\mathrm{E}_m - h\nu_1)t/\hbar} + \mathrm{F}_m{}^* e^{-i(\mathrm{E}_m + h\nu_1)t/\hbar}, \tag{11}$$

---

† (8) corresponds to "retarded" potentials. "Advanced" potentials, which we get from (8) by replacing *i* by − *i* in the exponential, would also give a solution of (4). It can be easily shown, however, that in the cases considered here, where at least one of the particles is treated as free, one gets the same result whether one uses retarded or advanced potentials or any combination of them with coefficients whose sum is equal to one. Hence in our case we could also replace the exponential in (8) by a cosine. (*Cf.* II.)

where ϕ_m is an eigenfunction of H and F_m and $F_m{}^*$ are independent of t. Introducing (11) into (9) and neglecting higher terms of U_1 and $U_1{}^*$ we get the following equations for F_m and $F_m{}^*$:

$$\{E_m - h\nu_1 - H\} F_m = - U_1 \phi_m \tag{12}$$

$$\{E_m + h\nu_1 - H\} F_m{}^* = - U_1{}^* \phi_m. \tag{12A}$$

These equations can be solved in the usual way. We expand the function F_m in terms of the complete set of eigenfunctions ϕ_s of the operator H: that is

$$F_m = \sum_s F_{ms} \phi_s.$$

Introducing this into (12) we get for the coefficients in the ordinary way

$$F_{ms} = - \frac{(s \mid U_1 \mid m)}{E_m - E_s - h\nu_1},$$

where $(s \mid U_1 \mid m)$ is a matrix element of the operator U_1. So we get for F_m the expression

$$F_m = - \sum_s \frac{(s \mid U_1 \mid m)}{E_m - E_s - h\nu_1} \phi_s. \tag{13}$$

As (12A) is obtained from (12) in replacing ν_1 and U_1 by $-\nu_1$ and $U_1{}^*$ respectively we get for $F_m{}^*$

$$F_m{}^* = - \sum_s \frac{(s \mid U_1{}^* \mid m)}{E_m - E_s + h\nu_1} \phi_s,$$

or as we have from (10) $\overline{(s \mid U_1{}^* \mid m)} = (m \mid U_1 \mid s)$

$$\bar{F}_m{}^* = - \sum_s \frac{(m \mid U_1 \mid s)}{E_m - E_s + h\nu_1} \bar{\phi}_s. \tag{13A}$$

The general solution ψ of (9) can now be obtained by superposition of all possible solutions χ_m; that is

$$\psi = \sum_m \chi_m$$

(for the moment we do not suppose the χ's to be normalized). For the corresponding current density we get neglecting higher terms in F and F* and arranging the sums in a suitable way†

$$\mathbf{J} = ce \sum_{ms} \bar{\chi}_m \alpha \chi_s = \mathbf{J}^0 + \mathbf{J}^1, \tag{14}$$

---

† *Cf.* Klein and Nishina, 'Z. Physik,' vol. 52, p. 860 (1929).

where $\mathbf{J}^0$ is the current density corresponding to the undisturbed eigenfunctions and

$$\mathbf{J}^1 = ce \sum_{ml} (\bar{\phi}_m \alpha \mathrm{F}_l + \bar{\mathrm{F}}_m{}^* \alpha \phi_l)\, e^{i(\mathrm{E}_m - \mathrm{E}_e + h\nu_1)t/\hbar} + \text{(conjugate complex term).}$$

Let us consider the term for which $l = n_2{}^0$ and $m = n_2$. Writing the suffix 2 in again, and using (6), (13), (13A), we get for this term of $\mathbf{J}^1$

$$\mathbf{J}_2 = \mathbf{J}'_2\, e^{-2\pi i \nu t} + \text{(conjugate complex)} \tag{15}$$

with

$$h\nu = \mathrm{E}_1{}^0 + \mathrm{E}_2{}^0 - \mathrm{E}_1 - \mathrm{E}_2 = k \tag{15A}$$

and

$$\mathbf{J}'_2 = -\, ce_2 \sum_{s_2} \left\{ \frac{\bar{\phi}_{n_2} \alpha \phi_{s_2} (s_2 \,|\, \mathrm{U}_1 \,|\, n_2{}^0)}{\mathrm{E}_2{}^0 - \mathrm{E}_{s_2} - \mathrm{E}_1 + \mathrm{E}_1{}^0} + \frac{(n_2 \,|\, \mathrm{U}_1 \,|\, s_2)\, \bar{\phi}_{s_2} \alpha \phi_{n_2}}{\mathrm{E}_2 - \mathrm{E}_{s_2} + \mathrm{E}_1 - \mathrm{E}_1{}^0} \right\}. \tag{16}$$

Supposing the quantity (15A) to be positive, it is clear that the current (15), which oscillates with the frequency ν, must be connected with the radiation emitted during the transition $(n_1{}^0 n_2{}^0) \rightarrow (n_1 n_2)$. We cannot, however, directly take (15) as the current producing this radiation, as it refers only to one particle. Until now the method has been unsymmetrical with respect to the two particles. It is clear that there must be complete symmetry. Simultaneously with (3) we have, therefore, to consider the zero order charge and current density corresponding to a transition of particle 2 between the states $n_2{}^0$ and n_2. We get them from (3) in replacing the suffix 1 by the suffix 2. In the same way we get the corresponding potentials from (5), (7), or (8). Then we consider again to the first order the change in the current density of the first particle which arises from these potentials and pick out the term which oscillates with the frequency (15A). This term we get from (15) and (16) by interchanging the suffixes 1 and 2; that is

$$\mathbf{J}_1 = \mathbf{J}'_1\, e^{-2\pi i \nu t} + \text{(conjugate complex)} \tag{15A}$$

with

$$\mathbf{J}'_1 = -\, ce_1 \sum_{s_1} \left\{ \frac{\bar{\phi}_{n_1} \alpha \phi_{s_1} (s_1 \,|\, \mathrm{U}_2 \,|\, n_1{}^0)}{\mathrm{E}_1{}^0 - \mathrm{E}_{s_1} - \mathrm{E}_2 + \mathrm{E}_2{}^0} + \frac{(n_1 \,|\, \mathrm{U}_2 \,|\, s_1)\, \bar{\phi}_{s_1} \alpha \phi_{n_1}}{\mathrm{E}_1 - \mathrm{E}_{s_1} + \mathrm{E}_2 - \mathrm{E}_2{}^0} \right\}. \tag{16A}$$

The whole current oscillating wth the frequency ν is then given by

$$\mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2 = \mathbf{J}' e^{-2\pi i \nu t} + \text{(conjugate complex)} \tag{17}$$

with

$$\mathbf{J}' = \mathbf{J}'_1 + \mathbf{J}'_2. \tag{18}$$

(All functions ϕ are now assumed to be normalized.)

We now have to calculate the field produced by the current density (17) according to Maxwell's equations. The solution is given in the usual way by means of retarded potentials. Let P be a point lying in the direction $\mathbf{n}$ and at a distance r from a fixed point inside the region (the box) where the collision takes place. If r is large compared with the dimensions of this region we get for the vector potential at P

$$\mathbf{A} = \frac{1}{cr} e^{-2\pi i\nu(t-r/c)} \mathbf{B} + \text{(conjugate complex)} \tag{19}$$

with

$$\mathbf{B} = \int \mathbf{J}' e^{-2\pi i\nu(\mathbf{n},\mathbf{x})/c} \, d\mathbf{x} = \int \mathbf{J}' e^{-i(\mathbf{k},\mathbf{x})/\hbar c} \, d\mathbf{x} \tag{20}$$

$$\mathbf{k} = k\mathbf{n} = h\nu\mathbf{n}. \tag{21}$$

The integration in (20) is to be extended over the whole space.

Equation (19) represents an outgoing electromagnetic wave. The intensity at the point P of the radiation component corresponding to a definite direction of polarization is given by Poynting's theorem; we get the expression

$$\frac{c}{4\pi} 2 \left(\frac{2\pi\nu}{c^2}\right)^2 |\mathrm{B}_k|^2 \frac{1}{r^2},$$

where B_k is the component of $\mathbf{B}$ in the direction of polarization. This quantity multiplied by $r^2 d\Omega_k$ and divided by $h\nu$ will then give us the probability that a light quantum of definite polarization and energy $k = h\nu$ is emitted into the solid angle $d\Omega_k$ during the transition

$$(n_1{}^0 n_2{}^0) \to (n_1 n_2).$$

Using (20), (18), (16), and (16A) we get for this probability

$$\begin{aligned}\Phi_{n_1{}^0 n_2{}^0, n_1 n_2} = \frac{2\pi\nu}{hc^3} d\Omega_k \Bigg| - ce_2 \sum_{s_2} \Bigg\{ & \frac{(n_2 \,|\, \alpha_k e^{-i(\mathbf{k},\mathbf{x}_2)/\hbar c} \,|\, s_2)(s_2 \,|\, \mathrm{U}_1 \,|\, n_2{}^0)}{\mathrm{E}_2{}^0 - \mathrm{E}_{s_2} - \mathrm{E}_1 + \mathrm{E}_1{}^0} \\ & + \frac{(n_2 \,|\, \mathrm{U}_1 \,|\, s_2)(s_2 \,|\, \alpha_k e^{-i(\mathbf{k},\mathbf{x}_2)/\hbar c} \,|\, n_2{}^0)}{\mathrm{E}_2 - \mathrm{E}_{s_2} + \mathrm{E}_1 - \mathrm{E}_1{}^0} \Bigg\} + \begin{pmatrix}\text{term with 1 and}\\ \text{2 interchanged}\end{pmatrix} \Bigg|^2 \\ = \frac{2\pi\nu}{hc^3} d\Omega_k & \left| \mathrm{X}(n_1, n_2;\; n_1{}^0, n_2{}^0) \right|^2, \end{aligned} \tag{22}$$

where α_k is the component of α in the direction of polarization.

If the two particles are electrons one must take the exclusion principle into account. To do this one has to replace the quantity

$$\mathrm{X}(n_1, n_2;\; n_1{}^0, n_2{}^0),$$

inside $|\quad|^2$ in (22) by $X(n_1, n_2; n_1^0, n_2^0) - X(n_2, n_1; n_1^0, n_2^0)$; that is

$$\Phi_{n_1^0 n_2^0, n_1 n_2} = \frac{2\pi\nu}{hc^3}\, d\Omega_k \left| X(n_1, n_2;\ n_1^0, n_2^0) - X(n_2, n_1;\ n_1^0, n_2^0) \right|^2. \tag{22A}$$

Since

$$X(n_2, n_1;\ n_1^0, n_2^0) = X(n_1, n_2;\ n_2^0, n_1^0),$$

(22A) vanishes if either $n_1 = n_2$ or $n_1^0 = n_2^0$ in accordance with the exclusion principle.

Let us now consider a little more closely the case of free particles. We introduce the special expressions for the ϕ's and for U_1 given by (1), (2), (10), and (7). Since there is " conservation of momentum " in all the matrix elements appearing in (22) the infinite sums over s_2 reduce to finite sums over the four states of a definite momentum. Also one sees that only those processes are possible for which it is true that

$$\mathbf{p}_1^0 + \mathbf{p}_2^0 = \mathbf{p}_1 + \mathbf{p}_2 + \mathbf{k}. \tag{23}$$

We now want to calculate the cross-section $d\Phi$ for the following process: the particle 1 is scattered into the solid angle $d\Omega_{E_1}$, having an energy between E_1 and $E_1 + dE_1$ after the collision; a light quantum is emitted into the solid angle $d\Omega_k$. To obtain this cross-section we have to sum (22) over all values of $\mathbf{p}_2$ and over all values of $\mathbf{p}_1$ lying inside the region $[d\Omega_{E_1};\ (E_1,\ E_1 + dE_1)]$, the number of these $\mathbf{p}_1$-values being equal to $d\Omega_{E_1} p_1 E_1\, dE_1 V/h^3c^3$. According to the method of normalization adopted in (2) we have further to divide by $v_1^0/V = cp_1^0/E_1^0V$ (v_1^0 is the initial velocity of particle 1).

In this way we get

$$\begin{aligned} d\Phi = d\Omega_k\, d\Omega_{E_1} \frac{2\pi}{hc} \frac{p_1 E_1 E_1^0 k\, dE_1}{\pi^2 p_1^0} \cdot \Bigg| & \frac{e_1^2 e_2}{|\mathbf{p}_2 - \mathbf{p}_2^0| - (E_2 - E_2^0)^2} \\ & \times \left\{ \sum_{1'} \frac{\bar{u}_1 \alpha_k u'_1 \left[\bar{u}'_1 u_1^0 \cdot \bar{u}_2 u_2^0 - (\bar{u}'_1 \alpha u_1^0, \bar{u}_2 \alpha u_2^0)\right]}{E_1^0 - E'_1 - E_2 + E_2^0} \right. \\ & \left. + \sum_{1''} \frac{\left[\bar{u}_2 u_2^0 \cdot \bar{u}_1 u''_1 - (\bar{u}_2 \alpha u_2^0, \bar{u}_1 \alpha u''_1)\right] \bar{u}''_1 \alpha_k u_1^0}{E_1 - E''_1 + E_2 - E_2^0} \right\} \\ & + (1 \text{ and } 2 \text{ interchanged}) \Bigg|^2. \end{aligned} \tag{24}$$

In this expression the values of $\mathbf{p}_2$ and k are understood to be given by the equations (23) and (15A)

$$k = E_1^0 - E_1 + E_2^0 - E_2.$$

The sums $\sum_{1'}$ and $\sum_{1''}$ appearing in (24) are to be extended over the four states corresponding to the momenta

$$\mathbf{p}_1' = \mathbf{p}_1 + \mathbf{k},\ \mathbf{p}_1'' = \mathbf{p}_1^0 - \mathbf{k},$$

respectively. If the particles are electrons we have inside $|\quad|^2$ in (24) to subtract the term which one gets by interchanging the final states $(\mathbf{p}_1 \mathrm{E}_1)$ and $(\mathbf{p}_2 \mathrm{E}_2)$ (*cf.* (22A)).

Let us now consider the case where one of the particles, say particle 2, becomes infinitely heavy. For $m_2 \rightarrow \infty$ with finite momenta we have $\mathrm{E}_2 - \mathrm{E}_2^0 \rightarrow 0$ and $\overline{u}'_2 \alpha u''_2 \rightarrow 0$, where u'_2 and u''_2 here are the amplitudes of two arbitrary waves. Therefore the term (1 and 2 interchanged) goes to zero. We have also $\overline{u}_2 u_2^0 \rightarrow 1$ or 0 according to the amplitudes u_2 and u_2^0 correspond to the same or to different directions of spin. If we put $e_2 = Ze$ and $e_1 = e$ and drop the suffix 1 we get then for $d\Phi$ in this limiting case

$$d\Phi = d\Omega_k\, d\Omega_{\mathrm{E}} \frac{2\pi e^2}{hc} \frac{Ze^4}{\pi^2} \frac{p\mathrm{E}\mathrm{E}^0 k\, d\mathrm{E}}{p^0 q^4} \left| \Sigma \frac{\overline{u}\alpha_k u' \,.\, \overline{u}' u^0}{\mathrm{E}^0 - \mathrm{E}'} + \Sigma \frac{\overline{u} u'' \,.\, \overline{u}'' \alpha_k u^0}{\mathrm{E} - \mathrm{E}''} \right|, \quad (25)$$

with

$$q = |\mathbf{p}_2 - \mathbf{p}_2^0| = |\mathbf{p}^0 - \mathbf{p} - \mathbf{k}| \qquad \text{and} \qquad k = \mathrm{E}^0 - \mathrm{E}.$$

(25) is just the formula of Bethe and Heitler.† It could be obtained by a direct application of Klein's correspondence method, since the field of the infinitely heavy particle can be treated as an external field.

The conditions that (24) shall go over into (25) in the limiting case $m_2 \rightarrow \infty$ removes any arbitrariness in our method. For instance, it is clear that we have to consider the real charge and current densities $\rho_{n_1 n_1^0} + \rho_{n_1^0 n_1}$, $\mathbf{J}_{n_1 n_1^0} + \mathbf{J}_{n_1^0 n_1}$ in (3) instead of just $\rho_{n_1 n_1^0}$ and $\mathbf{J}_{n_1 n_1^0}$, otherwise we should have got only the first term in (25).

Since we have used only relativistically invariant equations and expressions such as Dirac's equation, Dirac's expressions for charge and current density, and Maxwell's equations, it is clear that the method is relativistically invariant.

2—In this section we shall treat the problem by means of the quantum electrodynamics of Heisenberg, Pauli, and Dirac. Neglecting infinite terms arising from the self-energy of the particles, we get the same results as in § 1. We use the form which has recently been given to the theory of Fock.‡ According to Fock the behaviour of the system of particles

† 'Proc. Roy. Soc.,' A, vol. 146, p. 85 (1934).

‡ 'Phys. Z. Sowjet,' vol. 6, p. 425 (1934).

and photons can be described by a set of functions $\psi_0, \psi_1, \psi_2, \ldots, \psi_q, \ldots$, each of which is a function of the co-ordinates $\mathbf{x}_1$ and $\mathbf{x}_2$ of the particles. The q'th function ψ_q further depends on the momenta $\mathbf{k}_1/c, \mathbf{k}_2/c, \ldots, \mathbf{k}_q/c$ and directions of polarization $j_1, j_2, \ldots, j_q$ of q photons. These functions have the following physical meaning:

$$|\psi_q(\mathbf{x}_1, \mathbf{x}_2; \mathbf{k}_1 j_1, \mathbf{k}_2 j_2, \ldots, \mathbf{k}_q j_q)|^2,$$

and this expression gives the probability of finding one photon in each of the q states, $\mathbf{k}_1 j_1, \mathbf{k}_2 j_2, \ldots, \mathbf{k}_q j_q$ and the two particles at the places $\mathbf{x}_1$ and $\mathbf{x}_2$.

The functions ψ_q are determined by the following sequence of equations (*cf.* V. Fock, *loc. cit.*, equations (53), (54), and (71)):

$$\begin{aligned}\left\{\mathrm{H} - i\hbar\frac{\partial}{\partial t}\right\} \psi_q(\mathbf{k}_1, \mathbf{k}_2, \ldots \mathbf{k}_q) = &- \mathrm{V}\psi_q(\mathbf{k}_1, \ldots \mathbf{k}_q)\\ &+ \frac{\sqrt{q+1}}{(\hbar c)^3}\int d\mathbf{k}'\,\overline{\mathrm{G}}(\mathbf{k}')\,e^{-ik't/\hbar}\psi_{q+1}(\mathbf{k}', \mathbf{k}_1, \mathbf{k}_2, \ldots \mathbf{k}_q),\\ &+ \frac{1}{\sqrt{q}}\mathrm{G}(\mathbf{k}_1)\,e^{ik_1t/\hbar}\psi_{q-1}(\mathbf{k}_2, \ldots \mathbf{k}_q)\\ &+ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ &+ \frac{1}{\sqrt{q}}\mathrm{G}(\mathbf{k}_q)\,e^{ik_qt/\hbar}\psi_{q-1}(\mathbf{k}_1, \mathbf{k}_2, \ldots \mathbf{k}_{q-1}).\\ &(q = 0, 1, 2, \ldots).\end{aligned} \tag{26}$$

For simplicity we have dropped the arguments $\mathbf{x}_1$, $\mathbf{x}_2$, and the indices j in the ψ-functions. From now on the symbol $\mathbf{k}$ indicates a state in which the photon has the momentum $\mathbf{k}/c$ and a definite polarization. Accordingly the integration over $\mathbf{k}'$ in (26) is understood to include summation over the two directions of polarization. $\mathrm{H} = \mathrm{H}_1 + \mathrm{H}_2$ is the hamiltonian of the particles neglecting the interaction with the field: $\mathrm{G}(\mathbf{k})$ is given by

$$\left.\begin{aligned}\mathrm{G}(k) &= \sum_{s=1}^{2}\mathrm{G}_s(k)\\ \mathrm{G}_s(k) &= \frac{\hbar c}{2\pi}\frac{e_s}{\sqrt{k}}\,\alpha_k^{(s)}e^{-i(\mathbf{k}, \mathbf{x}_s)/\hbar c}\end{aligned}\right\}, \tag{27}$$

where $\alpha_k^{(s)}$ is the component in the direction of polarization of the Dirac matrix vector $\boldsymbol{\alpha}^{(s)}$ belonging to the sth particle. Finally $\overline{\mathrm{G}}$ is the conjugate operator to G and $\mathrm{V} = e_1e_2/|\mathbf{x}_1 - \mathbf{x}_2|$ is the electrostatic interaction between the particles. Actually we have already in (26) left out certain divergent integrals which correspond to the electrostatic self-energy of the particles.

To solve the equations (26) we expand the functions ψ_q in terms of the complete set of eigenfunctions ϕ_n of the operator H; that is

$$\psi_q(\mathbf{x}_1, \mathbf{x}_2; \mathbf{k}_1, \dots \mathbf{k}_q) = \sum_n a_q(n; \mathbf{k}_1, \dots \mathbf{k}_q)\, \phi_n(\mathbf{x}_1, \mathbf{x}_2)\, e^{-iE_n t/\hbar} \tag{28}$$
$$(q = 0, 1, 2, \dots).$$

Further we expand the coefficients a_q in terms of ascending powers of the charges e_1 and e_2; say

$$a_q(n; \mathbf{k}_1, \dots \mathbf{k}_q) = \sum_{\nu=0}^{\infty} a_q^{\nu}(n; \mathbf{k}_1, \dots \mathbf{k}_q), \tag{29}$$

where a_q^{ν} is of the ν'th order in the charges.

Since G is of the first and V is of the second order in the charges we get from (26), (28), and (29) the following equations for the determination of the coefficients a_q^{ν}

$$\begin{aligned}\frac{\hbar}{i}\frac{d}{dt} a_q^{\nu}(n; \mathbf{k}_1, \dots \mathbf{k}_q) = &- (n\,|\,\mathrm{V}\,|\,m)\, a_q^{\nu-2}(m; \mathbf{k}_1, \dots \mathbf{k}_q)\, e^{i(E_n - E_m)t/\hbar} \\ &+ \frac{\sqrt{q+1}}{(\hbar c)^3}\int d\mathbf{k}'\,(n\,|\,\overline{\mathrm{G}}(\mathbf{k}')\,|\,m)\, a_{q+1}^{\nu-1}(m; \mathbf{k}', \mathbf{k}_1, \dots \mathbf{k}_q)\, e^{i(E_n - E_m - k')t/\hbar} \\ &+ \frac{1}{\sqrt{q}}(n\,|\,\mathrm{G}(\mathbf{k}_1)\,|\,m)\, a_{q-1}^{\nu-1}(m; \mathbf{k}_2, \dots \mathbf{k}_q)\, e^{i(E_n - E_m + k_1)t/\hbar} \\ &\quad + \dots\dots\dots\dots\dots\dots\dots\dots \\ &+ \frac{1}{\sqrt{q}}(n\,|\,\mathrm{G}(\mathbf{k}_q)\,|\,m)\, a_{q-1}^{\nu-1}(m; \mathbf{k}_1, \dots \mathbf{k}_{q-1})\, e^{i(E_n - E_m + k_q)t/\hbar}\end{aligned} \tag{30}$$

for $(q, \nu = 0, 1, 2, \dots)$ independently.

We have here made the convention often made in tensor calculus that indices like m in (30), which appear twice, have to be summed over.

Let us assume that there are no photons present at the time $t = 0$ and that the system of the two particles is in the state n^0; that is

for $t = 0$

and arbitrary ν

and

$$\left.\begin{aligned} a_q^{\nu} &= 0 \quad \text{for} \quad q \geqslant 1 \\ a_0^{\nu} &= 0 \quad \text{for} \quad \nu \geqslant 1 \\ a_0^{0}(n) &= \delta_{nn^0}. \end{aligned}\right\} \tag{31}$$

Since from (30) we have

we have

for all times.

$$\left.\begin{aligned} \frac{da_0^0}{dt} &= 0 \\ a_0^0(n) &= \delta_{n^0 n} \end{aligned}\right\} \tag{32}$$

Using (31) and (30) one finds easily by induction that

$$a_q^{\nu} = 0 \quad \text{for} \quad q \geqslant \nu + 1 \text{ and arbitrary } \nu \tag{33}$$

and also

$$a_q^{q+2\mu-1} = 0 \text{ for } (q, \mu) = 0, 1, 2, \ldots \text{ independently.}$$

From (30), (31), and (32) we get

$$\begin{aligned}\frac{\hbar}{i}\frac{d}{dt} a_1^{1}(n; \mathbf{k}') &= (n \,|\, \mathrm{G}(\mathbf{k}') \,|\, m)\, \delta_{n^0 m}\, e^{i(\mathrm{E}_n - \mathrm{E}_m + k')t/\hbar} \\ &= (n \,|\, \mathrm{G}(\mathbf{k}') \,|\, n^0)\, e^{i(\mathrm{E}_n - \mathrm{E}_{n^0} + k')t/\hbar}\end{aligned}$$

and

$$a_1^{1}(n; \mathbf{k}') = (n \,|\, \mathrm{G}(\mathbf{k}') \,|\, n^0)\, \frac{e^{i(\mathrm{E}_n - \mathrm{E}_{n^0} + k')t/\hbar} - 1}{\mathrm{E}_n - \mathrm{E}_{n^0} + k'}\,. \tag{34}$$

The radiation corresponding to this first approximation is equal to that which one gets from the "unperturbed" current density J^0 in (14) and will not concern us here. In the case of free particles it gives no contribution at all.

Using (30), (31), (32), and (33) we get

$$\begin{aligned}\frac{\hbar}{i}\frac{d}{dt} a_0^{2}(n) = &- (n \,|\, \mathrm{V} \,|\, n^0)\, e^{i(\mathrm{E}_n - \mathrm{E}_{n^0})t/\hbar} \\ &+ \frac{1}{(\hbar c)^3}\int d\mathbf{k}'\, (n \,|\, \overline{\mathrm{G}}(\mathbf{k}') \,|\, m)\, (m \,|\, \mathrm{G}(\mathbf{k}') \,|\, n^0)\, \frac{e^{i(\mathrm{E}_m - \mathrm{E}_0 + k')t/\hbar} - 1}{\mathrm{E}_m - \mathrm{E}_{n^0} + k'} \,.\, e^{i(\mathrm{E}_n - \mathrm{E}_m - k')t/\hbar}\end{aligned}$$

and

$$\begin{aligned}a_0^{2}(n) = &\left\{ - (n \,|\, \mathrm{V} \,|\, n^0) + \frac{1}{(\hbar c)^3}\int d\mathbf{k}'\, \frac{(n \,|\, \overline{\mathrm{G}}(\mathbf{k}') \,|\, m)\, (m \,|\, \mathrm{G}(\mathbf{k}') \,|\, n^0)}{\mathrm{E}_m - \mathrm{E}_{n^0} + k'} \right\} \\ &\times \frac{e^{i(\mathrm{E}_n - \mathrm{E}_{n^0})t/\hbar} - 1}{\mathrm{E}_n - \mathrm{E}_{n^0}} - \frac{1}{(\hbar c)^3}\int dk'\, \frac{(n \,|\, \overline{\mathrm{G}}(k') \,|\, m)\, (m \,|\, \mathrm{G}(k') \,|\, n^0)}{\mathrm{E}_m - \mathrm{E}_{n^0} + k'} \\ &\cdot \frac{e^{i(\mathrm{E}_n - \mathrm{E}_m - k')t/\hbar} - 1}{\mathrm{E}_n - \mathrm{E}_m - k'}\,.\end{aligned} \tag{35}$$

This approximation gives, apart from certain infinite terms, the formula for the non-radiative collision derived by the correspondence method in I.†

Using again (30), (31), (33), and (34), we have

$$\begin{aligned}\frac{\hbar}{i}\frac{d}{dt} a_2^{2}(n; \mathbf{k}', \mathbf{k}) = &\frac{1}{\sqrt{2}}\, (n \,|\, \mathrm{G}(\mathbf{k}') \,|\, m)\, (m \,|\, \mathrm{G}(\mathbf{k}) \,|\, n^0)\, \frac{e^{i(\mathrm{E}_m - \mathrm{E}_{n^0} + k)t/\hbar} - 1}{\mathrm{E}_m - \mathrm{E}_{n^0} + k} \\ &.\, e^{i(\mathrm{E}_n - \mathrm{E}_m + k')t/\hbar} + (\text{term with } \mathbf{k} \text{ and } \mathbf{k}' \text{ interchanged})\end{aligned}$$

---

† *Cf.* V. Fock, 'Phys. Z. Sowjet,' vol. 6, p. 463 (1934); Fock and Podolsky, 'Phys. Z. Sowjet,' vol. 2, p. 275 (1932); Bethe and Fermi, 'Z. Physik,' vol. 77, p. 296 (1932).

and

$$a_2^2(n;\mathbf{k}',\mathbf{k}) = \frac{1}{\sqrt{2}}\frac{(n\,|\,\mathrm{G}(\mathbf{k}')\,|\,m)\,(m\,|\,\mathrm{G}(\mathbf{k})\,|\,n^0)}{\mathrm{E}_m-\mathrm{E}_{n^0}+k}$$
$$\times\left\{\frac{e^{i(\mathrm{E}_n-\mathrm{E}_{n^0}+k+k')t/\hbar}-1}{\mathrm{E}_n-\mathrm{E}_{n_0}+k+k'}-\frac{e^{i(\mathrm{E}_n-\mathrm{E}_m+k')t/\hbar}-1}{\mathrm{E}_n-\mathrm{E}_m+k'}\right\}$$
$$+\ (\text{term with } \mathbf{k} \text{ and } \mathbf{k}' \text{ interchanged}). \tag{36}$$

Finally we get from (30)

$$\frac{\hbar}{i}\frac{d}{dt}a_1^3(n;\mathbf{k}) = -(n\,|\,\mathrm{V}\,|\,m)\,a_1^1(m;\mathbf{k})\,e^{i(\mathrm{E}_n-\mathrm{E}_m)t/\hbar}$$
$$+\frac{\sqrt{2}}{(\hbar c)^3}\int d\mathbf{k}'\,(n\,|\,\bar{\mathrm{G}}(\mathbf{k}')\,|\,m)\,a_2^2(m;\mathbf{k}',\mathbf{k})\,e^{i(\mathrm{E}_n-\mathrm{E}_m-k')t/\hbar}$$
$$+(n\,|\,\mathrm{G}(\mathbf{k})\,|\,m)\,a_0^2(m)\,e^{i(\mathrm{E}_n-\mathrm{E}_m+k)t/\hbar}.$$

Introducing the expressions for a_1^1, a_2^2, and a_0^2 given by (34), (35), and (36) and integrating we get

$$a_1^3(n;\mathbf{k}) = \left\{\frac{(n\,|\,\mathrm{G}(\mathbf{k})\,|\,m)}{\mathrm{E}_m-\mathrm{E}_{n^0}}\cdot\Bigg[-(m\,|\,\mathrm{V}\,|\,n^0)\right.$$
$$+\frac{1}{(\hbar c)^3}\int d\mathbf{k}'\,\frac{(m\,|\,\bar{\mathrm{G}}(\mathbf{k}')\,|\,l)\,(l\,|\,\mathrm{G}(\mathbf{k}')\,|\,n^0)}{\mathrm{E}_l-\mathrm{E}_{n^0}+k'}\Bigg]+\Bigg[-(n\,|\,\mathrm{V}\,|\,m)$$
$$+\frac{1}{(\hbar c)^3}\int d\mathbf{k}'\,\frac{(n\,|\,\bar{\mathrm{G}}(\mathbf{k}')\,|\,l)\,(l\,|\,\mathrm{G}(\mathbf{k}')\,|\,m)}{\mathrm{E}_l-\mathrm{E}_{n^0}+k+k'}\Bigg]\frac{(m\,|\,\mathrm{G}(\mathbf{k})\,|\,n^0)}{\mathrm{E}_m-\mathrm{E}_{n^0}+k}$$
$$\left.+\frac{1}{(\hbar c)^3}\int d\mathbf{k}'\,\frac{(n\,|\,\bar{\mathrm{G}}(\mathbf{k}')\,|\,l)\,(l\,|\,\mathrm{G}(\mathbf{k})\,|\,m)\,(m\,|\,\mathrm{G}(\mathbf{k}')\,n^0)}{(\mathrm{E}_m-\mathrm{E}_{n^0}+k')(\mathrm{E}_l-\mathrm{E}_{n^0}+k+k')}\right\}$$
$$\times\frac{e^{i(\mathrm{E}_n-\mathrm{E}_{n^0}+k)t/\hbar}-1}{\mathrm{E}_n-\mathrm{E}_{n^0}+k}. \tag{37}$$

In this expression we have left out all those terms which do not fulfil the law of the conservation of energy.

Since $\mathrm{H} = \mathrm{H}_1 + \mathrm{H}_2$ is the sum of the hamiltonians of the two particles the eigenfunctions ϕ_n of H are products of the eigenfunctions ϕ_{n_1} and ϕ_{n_2} of H_1 and H_2 respectively.† Accordingly E_n is equal to $\mathrm{E}_{n_1} + \mathrm{E}_{n_2}$ and similarly for the other states appearing in (37). Introducing (27) into (37) one sees at once that only terms which are proportional to $e_1e_2^2$ and $e_2e_1^2$ appear. The terms proportional to e_1^3 and e_2^3 vanish

---

† If the particles are electrons one must take the combination

$$\phi_n = \frac{1}{\sqrt{2}}[\phi_{n_1}(\mathbf{x}_1)\,\phi_{n_2}(\mathbf{x}_2) - \phi_{n_2}(\mathbf{x}_1)\,\phi_{n_1}(\mathbf{x}_2)].$$

because of the orthogonality of the eigenfunctions $\phi_{n_1}, \phi_{n_2^0}$ and $\phi_{n_1}, \phi_{n_2^0}$. After a rather long calculation we get, therefore,

$$
\begin{aligned}
a_1^3(n_1, n_2; \mathbf{k}) = \Bigg\{ & \frac{(n_2 \mid G_2(\mathbf{k}) \mid m_2)}{E_{m_2} - E_{n_2^0} + E_{n_1} - E_{n_1^0}} \Bigg[- (n_1 m_2 \mid V \mid n_1^0 n_2^0) \\
& + \frac{1}{(\hbar c)^3} \int dk' \, \frac{2k' \, (m_2 \mid \overline{G}_2(\mathbf{k}') \mid n_2^0) \, . \, (n_1 \mid G_1(\mathbf{k}') \mid n_1^0)}{k'^2 - (E_{n_1} - E_{n_1^0})^2} \Bigg] \\
& + \Bigg[- (n_1 n_2 \mid V \mid n_1^0 m_2) \\
& + \frac{1}{(\hbar c)^3} \int dk' \, \frac{2k' \, (n_2 \mid \overline{G}_2(k') \mid m_2)(n_1 \mid G_1(\mathbf{k}') \mid n_1^0)}{k'^2 - (E_{n_1} - E_{n_1^0})^2} \Bigg] \\
& \times \frac{(m_2 \mid G_2(\mathbf{k}) \mid n_2^0)}{E_{m_2} - E_{n_2} + E_{n_1^0} - E_{n_1}} \\
& + \begin{pmatrix} \text{term with 1 and 2} \\ \text{interchanged} \end{pmatrix} \Bigg\} \frac{e^{i(E_{n_1} + E_{n_2} - E_{n_1^0} - E_{n_2^0} + k)t/\hbar} - 1}{E_{n_1} + E_{n_2} - E_{n_1^0} - E_{n_2^0} + k} .
\end{aligned} \tag{38}
$$

Here we have already made use of the law of conservation of energy which is secured by the time-factor outside the brackets. We now have to evaluate the terms in square brackets. To give a definite meaning to the improper integrals we take their principal values. We first perform the summation over the two directions of polarization implied in $\int d\mathbf{k}'$.

Now if Σ means this summation one has†

$$
\begin{aligned}
\Sigma \, \overline{G}_2(\mathbf{k}') \, G_1(\mathbf{k}') &= \left(\frac{\hbar c}{2\pi}\right)^2 \frac{e_1 e_2}{k'} e^{-i(\mathbf{k}', \mathbf{x}_1 - \mathbf{x}_2)/\hbar c} \, \Sigma \, \alpha_{k'}^{(1)} \alpha_{k'}^{(2)} \\
&= \left(\frac{\hbar c}{2\pi}\right)^2 \frac{e_1 e_2}{k'} e^{-i(\mathbf{k}', \mathbf{x}_1 - \mathbf{x}_2)/\hbar c} \left[(\alpha^{(1)}, \alpha^{(2)}) - \frac{1}{k'^2} (\alpha^{(1)}, \mathbf{k}')(\alpha^{(2)}, \mathbf{k}') \right] .
\end{aligned} \tag{39}
$$

Further, if Φ^0 and $\mathbf{A}^0$ are the potentials of the external field, we have

$$
H_1 = - c\,(\alpha^{(1)}, \mathbf{p}^{(1)}) + \rho_3^{(1)} m_1 c^2 - e_1 \Phi^0 - e_1 (\alpha^{(1)}, \mathbf{A}^0),
$$

and therefore

$$
(\alpha^{(1)}, \mathbf{k}') \, e^{-i(\mathbf{k}', \mathbf{x}_1)/\hbar c} = (\alpha^{(1)}, i\hbar c \, \mathrm{grad} \, e^{-i(\mathbf{k}', \mathbf{x}_1)/\hbar c}) = H_1 e^{-i(\mathbf{k}', \mathbf{x}_1)/\hbar c} - e^{-i(\mathbf{k}' \mathbf{x}_1)/\hbar c} H_1 .
$$

Similarly we get

$$
(\alpha^{(2)}, \mathbf{k}') \, e^{i(\mathbf{k}' \mathbf{x}_2)/\hbar c} = - [H_2 e^{i(\mathbf{k}', \mathbf{x}_2)/\hbar c} - e^{i(\mathbf{k}', \mathbf{x}_2)/\hbar c} H_2].
$$

---

† Fock, 'Phys. Z. Sowjet,' vol. 6, p. 425, equation (84) (1934).

So we have from (39)

$$(n_1 m_2 \mid \Sigma \,\overline{\mathrm{G}}_2(\mathbf{k}')\,\mathrm{G}_1(\mathbf{k}') \mid n_1^0 n_2^0)$$
$$= \left(\frac{\hbar c}{2\pi}\right)^2 \frac{e_1 e_2}{k'} \Big\{ (n_1 m_2 \mid (\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)})\, e^{i(\mathbf{k}', \mathbf{x}_1 - \mathbf{x}_2)/\hbar c} \mid n_1^0 n_2^0)$$
$$+ \frac{(\mathrm{E}_{n_1} - \mathrm{E}_{n_1^0})(\mathrm{E}_{m_2} - \mathrm{E}_{n_2^0})}{k'^2} (n_1 m_2 \mid e^{-i(\mathbf{k}', \mathbf{x}_1 - \mathbf{x}_2)/\hbar c} \mid n_1^0 n_2^0) \Big\}. \tag{40}$$

Using (40) and the formula

$$\mathrm{V} = \frac{e_1 e_2}{|x_1 - x_2|} = \frac{e_1 e_2}{2\pi^2 \hbar c} \int \frac{e^{-i(\mathbf{k}', \mathbf{x}_1 - \mathbf{x}_2)/\hbar c}}{k'^2}\, dk',$$

we can write the first square bracket in (38) in the following form:

$$\frac{e_1 e_2}{2\pi^2 \hbar c} \int dk' \frac{(n_1 m_2 \mid e^{-i(\mathbf{k}', \mathbf{x}_1 - \mathbf{x}_2)/\hbar c}\,[-1 + (\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)})] \mid n_1^0 n_2^0)}{k'^2 - (\mathrm{E}_{n_1} - \mathrm{E}_{n_1^0})^2}$$
$$+ \frac{e_1 e_2}{2\pi^2 \hbar c} (\mathrm{E}_{n_1} - \mathrm{E}_{n_1^0})(\mathrm{E}_{m_2} - \mathrm{E}_{n_2^0} + \mathrm{E}_{n_1} - \mathrm{E}_{n_1^0}) \int \frac{d\mathbf{k}'\,(n_1 m_2 \mid e^{-i(\mathbf{k}', \mathbf{x}_1 - \mathbf{x}_2)/\hbar c} \mid n_1^0 n_2^0)}{k'^2\,[k'^2 - (\mathrm{E}_{n_1} - \mathrm{E}_{n_1^0})^2]}. \tag{41}$$

Now we have†

$$\frac{1}{2\pi^2 \hbar c} \int d\mathbf{k}' \frac{e^{-i(\mathbf{k}', \mathbf{x}_1 - \mathbf{x}_2)/\hbar c}}{k'^2 - (\mathrm{E}_{n_1} - \mathrm{E}_{n_1^0})^2} = \frac{\cos\left(\dfrac{\mathrm{E}_{n_1} - \mathrm{E}_{n_1^0}}{\hbar c} |\mathbf{x}_1 - \mathbf{x}_2|\right)}{|x_1 - x_2|}.$$

Using (8) and (10) we see that the first term in (41) is just $(m_2 \mid \mathrm{U}_1 \mid n_2^0)$ (*cf.* also the footnote, p. 484). We get the second square bracket in (38) from (41) by replacing m_2, n_2^0 by n_2, m_2 respectively. Introducing this into (38) we get

$$a_1^3(n_1, n_2; \mathbf{k}) = \Big\{ \frac{e_1 e_2}{2\pi^2 \hbar c} (\mathrm{E}_{n_1} - \mathrm{E}_{n_1^0}) \int \frac{d\mathbf{k}'\,(n_1 \mid e^{i(\mathbf{k}'\mathbf{x}_1)/\hbar c} \mid n_1^0)}{k'^2\,[k'^2 - (\mathrm{E}_{n_1} - \mathrm{E}_{n_1^0})^2]}$$
$$\cdot\, (n_2 \mid \mathrm{G}_2(\mathbf{k})\, e^{i(\mathbf{k}'\mathbf{x}_2)/\hbar c} - e^{i(\mathbf{k}'\mathbf{x}_2)/\hbar c}\, \mathrm{G}_2(\mathbf{k}) \mid n_2^0)$$
$$+ \frac{\hbar c}{2\pi} \frac{e_2}{\sqrt{k}} \Big[\frac{(n_2 \mid \alpha_k^{(2)} e^{-i(\mathbf{k}, \mathbf{x}_2)/\hbar c} \mid m_2)(m_2 \mid \mathrm{U}_1 \mid n_2^0)}{\mathrm{E}_{m_2} - \mathrm{E}_{n_2^0} + \mathrm{E}_{n_1} - \mathrm{E}_{n_1^0}}$$
$$+ \frac{(n_2 \mid \mathrm{U}_1 \mid m_2)(m_2 \mid \alpha_k^{(2)} e^{-i(\mathbf{k}\mathbf{x}_2)/\hbar c} \mid n_2^0)}{\mathrm{E}_{m_2} - \mathrm{E}_{n_2} - \mathrm{E}_{n_1^0} + \mathrm{E}_{n_1}} \Big]$$
$$+ \left(\begin{matrix}\text{term with 1 and 2}\\ \text{interchanged}\end{matrix}\right\} \frac{e^{i(\mathrm{E}_{n_1} + \mathrm{E}_{n_2} - \mathrm{E}_{n_1^0} - \mathrm{E}_{n_2^0} + k)/\hbar} - 1}{\mathrm{E}_{n_1} + \mathrm{E}_{n_2} - \mathrm{E}_{n_1^0} - \mathrm{E}_{n_2^0} + k}. \tag{42}$$

Since the operators G_2 and $e^{i(\mathbf{k}'\mathbf{x}_2)/\hbar c}$ commute, the first term in (42) cancels. The probability per unit time for a transition in which the particles go over into the state $(n_1 n_2)$ and a photon is emitted with a momentum lying between $\mathbf{k}/c$ and $(\mathbf{k} + d\mathbf{k})/c$ is then given by

---

† *Cf.*, for instance, II, 'Ann. Physik, vol. 14, p. 543, equation (33).

$$\frac{d}{dt}\left| a_1^3 (n_1, n_2; \mathbf{k}) \right|^2 \frac{d\mathbf{k}.}{(\hbar c)^3}\,. \tag{43}$$

For large t the derivative with respect to time of the square of the time-factor in (42) becomes equal to $\frac{2\pi}{\hbar}\,\delta\,(\mathrm{E}_{n_1} + \mathrm{E}_{n_2} - \mathrm{E}_{n_1^0} - \mathrm{E}_{n_2^0} + k)$. Putting $d\mathbf{k} = k^2 dk\, d\Omega_k$ in (43) and integrating with respect to k we get for the probability of a transition $(n_1^0 n_2^0) \rightarrow (n_1 n_2)$ with emission of a photon into the solid angle $d\Omega_k$

$$\frac{2\pi k}{\hbar^2 c}\, d\,\Omega_k \left| e_2 \sum_{m_2} \left\{ \frac{(n_2 \,|\, \boldsymbol{\alpha}_k\, e^{-i(\mathbf{k}\mathbf{x}_2)/\hbar c} \,|\, m_2)\,(m_2 \,|\, \mathrm{U}_1 \,|\, n_2^0)}{\mathrm{E}_{m_2} - \mathrm{E}_{n_2^0} + \mathrm{E}_{n_1} - \mathrm{E}_{n_1^0}} + \frac{(n_2 \,|\, \mathrm{U}_1 \,|\, m_2)\,(m_2 \,|\, \boldsymbol{\alpha}_k\, e^{-i(\mathbf{k},\,\mathbf{x}_2)/\hbar c} \,|\, n_2^0)}{\mathrm{E}_{m_2} - \mathrm{E}_{n_2} + \mathrm{E}_{n_1^0} - \mathrm{E}_{n_1}} \right\} + \begin{pmatrix}\text{term with 1 and 2}\\ \text{interchanged}\end{pmatrix} \right|^2$$

with

$$k = h\nu = \mathrm{E}_{n_1^0} + \mathrm{E}_{n_2^0} - \mathrm{E}_{n_1} - \mathrm{E}_{n_2}.$$

Here we have again introduced the sign of summation.

This formula is in agreement with the formula (22) derived by the correspondence method in § 1. If we had taken antisymmetrical expressions for the eigenfunctions ϕ_n (*cf.* footnote, p. 493) we would have got the formula (22A).

This work was begun in Rome and finished in Cambridge. I should like to thank Professor E. Fermi and Professor R. H. Fowler for the very pleasant time spent at the institutes in Rome and Cambridge respectively. Further I should like to thank Dr. Hulme and Dr. Bhabha for many discussions about problems connected with this paper.

SUMMARY

In § 1 a correspondence method is developed for the treatment of the radiative collision between two particles. The method is relativistically invariant and forms an immediate generalization of the method previously used for the treatment of the non-radiative collision. In the limiting case, where one of the particles becomes infinitely heavy, the formula obtained for the differential cross-section goes over into the formula of Bethe and Heitler. In § 2 it is shown that the quantum electrodynamics of Heisenberg, Pauli, and Dirac, apart from infinite self-energy terms, gives the same result as the method used in § 1.

ERRATA.

Paper by J. C. Dobbie on "New terms in the spectrum of Fe II."
'Proc. Roy. Soc.,' A vol. 151, p. 703, October 1 (1935).

P. 703—Footnote †, vol. 6, No. 22 (1929) should *read* vol. 6, No. 11 (1929).

P. 711—Column 7, line 2, ·08 should *read* ·86.

P. 717—Column 4, 273·21 should *read* 273·31.

P. 718—Column 6, line 1, 0·80 should be *deleted.*

Plate 6, facing p. 724—"Spectrum of the iron arc (*a*) and (*b*) from . . ." should *read* "Spectrum of the iron arc (*a*) and spark (*b*) from . . ."

The Energies of Nuclear Reactions

By H. A. Wilson, F.R.S., Rice Institute, Houston, Texas

(*Received July* 12, 1935)

It has been shown in previous papers* that the nuclei of the radioactive atoms probably have equally spaced energy levels so that the energies of the rays emitted by the nucleus are equal to multiples of a constant $q = 3{\cdot}85 \times 10^5$ electron volts. Several light elements also appear to have nuclear levels with differences equal to q within the rather large limits of error. We should therefore expect the energies set free in nuclear reactions to be equal to multiples of q. The electronic energies of the light elements are small compared with q, so that the observed energies should be multiples of q.

The energies of a considerable number of nuclear reactions have been measured, but only a few with sufficient accuracy to show whether they are or are not equal to multiples of q. The energies of the 17 nuclear reactions given in a recent paper† by Oliphant, Kempton, and Rutherford are given in Table I. The energies are in atomic weight units equal to $9{\cdot}315 \times 10^8$ electron volts.‡ Some of the values have been increased slightly since they were not quite accurately calculated from the energies in electron volts.

TABLE I

$\mathrm{Li}^6 + \mathrm{H}^1 = \mathrm{He}^4 + \mathrm{He}^3 + 0{\cdot}00386$ (1)

$\mathrm{Li}^6 + \mathrm{H}^2 = 2\mathrm{He}^4 + 0{\cdot}02368$ ...... (2)

$\mathrm{Li}^6 + \mathrm{H}^2 = \mathrm{Li}^7 + \mathrm{H}^1 + 0{\cdot}00537$ (3)

$\mathrm{Li}^7 + \mathrm{H}^1 = 2\mathrm{He}^4 + 0{\cdot}01831$ ...... (4)

$\mathrm{Li}^7 + \mathrm{H}^2 = 2\mathrm{He}^4 + n^1 + 0{\cdot}0157$ (5)

$\mathrm{Be}^9 + \mathrm{H}^1 = \mathrm{Be}^8 + \mathrm{H}^2 + 0{\cdot}00051$ (6)

$\mathrm{Be}^9 + \mathrm{H}^1 = \mathrm{Li}^6 + \mathrm{He}^4 + 0{\cdot}0022$ (7)

$\mathrm{Be}^9 + \mathrm{H}^2 = \mathrm{Li}^7 + \mathrm{He}^4 + 0{\cdot}00775$ (8)

$\mathrm{Be}^9 + \mathrm{H}^2 = \mathrm{Be}^8 + \mathrm{H}^3 + 0{\cdot}0048$ (9)

$\mathrm{Be}^9 + \mathrm{H}^2 = \mathrm{Be}^{10} + \mathrm{H}^1 + 0{\cdot}0050$ (10)

$\mathrm{Be}^9 + \mathrm{H}^2 = \mathrm{B}^{10} + n^1 + 0{\cdot}0053$ (11)

$\mathrm{Be}^8 + n^1 = \mathrm{Be}^9 + 0{\cdot}0016$ ...... (12)

$\mathrm{B}^{11} + \mathrm{H}^1 = 3\mathrm{He}^4 + 0{\cdot}0091$ ...... (13)

$\mathrm{B}^{11} + \mathrm{H}^1 = \mathrm{Be}^8 + \mathrm{He}^4 + 0{\cdot}0092$ (14)

$\mathrm{B}^{10} + n^1 = \mathrm{Li}^7 + \mathrm{He}^4 + 0{\cdot}0021$ (15)

$\mathrm{H}^2 + \mathrm{H}^2 = \mathrm{He}^3 + n^1 + 0{\cdot}00285$ (16)

$\mathrm{H}^2 + \mathrm{H}^2 = \mathrm{H}^3 + \mathrm{H}^1 + 0{\cdot}00426$ (17)

The last figure in the energy values may be in error. The value of (5) is a lower limit. (9) is only roughly known.

These 17 equations are not all independent, so that it is possible to obtain several independent values of some of the energies and so get more probable values of them. Before comparing the energies with multiples of q we shall therefore revise the values of some of the less exactly known energies so as to make them consistent with the more exactly known values. In this way we can make the 17 equations consistent and improve the probable accuracy of some of the energies.

---

* 'Proc. Roy. Soc.,' A, vol. 150, p. 1 (1935).

† 'Proc. Roy. Soc.,' A, vol. 150, p. 241 (1935).

‡ $9{\cdot}315 \times 10^8 = (\text{velocity of light})^2/(\text{Faraday} \times 10^7)$.

There are six relations satisfied by the 17 equations, namely:

$$(2) = (3) + (4) \tag{A}$$
$$(8) = (7) + (3) \tag{B}$$
$$(8) = (11) + (15) \tag{C}$$
$$(9) = (6) + (17) \tag{D}$$
$$(4) = (5) + (6) + (12) \tag{E}$$
$$(2) + (6) + (13) + (16) = (1) + (5) + (8) + (14) \tag{F}$$

The equation (2) = (3) + (4), for example, means that if we add the two equations

$$\mathrm{Li}^6 + \mathrm{H}^2 = \mathrm{Li}^7 + \mathrm{H}^1 + 0\cdot00537 \tag{3}$$

and

$$\mathrm{Li}^7 + \mathrm{H}^1 = 2\mathrm{He}^4 + 0\cdot01831 \tag{4}$$

we get the equation

$$\mathrm{Li}^6 + \mathrm{H}^2 = 2\mathrm{He}^4 + 0\cdot02368. \tag{2}$$

The energies in Table I satisfy (2) = (3) + (4) but do not exactly satisfy the other five relations; for example, (11) + (15) = 0·0074, but (8) = 0·00775. It is convenient, for example, to use (8) to denote either the equation (8) or the energy of the reaction (8).

The equations (B), (C), and (F) give for the energy of (8) the values 0·00757, 0·0074, and 0·00738 respectively. Bethe's atomic weights* give (8) = 0·00743, and Oliphant, Kempton, and Rutherford's atomic weights† give (8) = 0·0076. The mean of these values is 0·00748. It seems probable, therefore, that the value 0·00775 given in Table I is too high, and we shall adopt 0·00748 as the probable value of the energy of (8). The equation (B), since (3) is known accurately, then requires (7) = 0·00211 instead of 0·0022. The equation (C) shows that (11) + (15) is too low, so we increase it in the ratio 740 to 748, making (11) = 0·00536 and (15) = 0·00212.

The value of (6) can be easily shown to be given by

$$(6) = \tfrac{1}{4}\{(7) - (12) + (1) - (16) + (14) - (13)\}$$

which gives (6) = 0·00081, which is much greater than the observed value 0·00051. The value of (7) is reliable but (12) = 0·0016 is probably too small. The value of (12), given by Aston's atomic weights and also by Oliphant, Kempton, and Rutherford's atomic weights, is 0·0017. We therefore adopt 0·0017 as the probable value of (12). This reduces the value of (6) given by the above expression to 0·00076, which is still too large. The difference (1) — (16) is probably reliable, so that (14) — (13) must be too large. To get (6) = 0·00051 it is necessary to change (14) — (13) from 0·0001 to —0·00040. We therefore change (14) to (13) — 0·00040 = 0·0087.

---

* 'Phys. Rev.,' vol. 47, p. 634 (1935).

† *Loc. cit.*

With the values adopted the equations (E) and (F) both give (5) = 0·0161. The value (5) = 0·0157 is a lower limit, so we adopt 0·0161 as the value of (5).

The equation (D) requires (9) = 0·00477 instead of 0·0048. The values (6) = 0·00051, (17) = 0·00426, and (9) = 0·00477 satisfy (D) but, as will be seen later, it is probable that (6) and (17) are both slightly too high, which makes (9) appreciably too high.

The revised energies satisfy the six relations between the 17 equations, so that with the revised values the equations are consistent.

The value $q = 3 \cdot 85 \times 10^5$ electron volts was based on the observed energies of beta and alpha rays. Scott* has recently made a very careful determination of the energy of the most intense beta ray from radium B and finds that the previous value is about ½% too low. This probably means that q should be increased by about ½%, and so is equal to $3 \cdot 87 \times 10^6$ electron volts. This higher value is used here, but it makes very little difference which value is used. The energy $3 \cdot 87 \times 10^6$ electron volts is equal to $q_m = 0 \cdot 000415$ in atomic weight units.

The energies of the 17 reactions are given in Table II. The revised energies are indicated by an asterisk. The quotients got by dividing the energies by $q_m = 0 \cdot 00415$ are also given, and the values of nq_m where n is the integer nearest to the quotient of the energy by q_m.

Among the differences between the quotients and integers there are nine between 0 and 0·1, four between 0·1 and 0·2, three between 0·2 and 0·3, and one equal to 0·49. The quotient 11·49 for reaction (9) is of very little significance because the energy of (9) is only very roughly known and may well be equal to $11q = 0 \cdot 004565$ instead of 0·00477.

The chance that the approximate agreements of the energies with the multiples of q_m is due to accidental coincidences is clearly very small. Disregarding the energy of (9), it can be easily shown that the chance of getting no quotients differing from integers by more than 0·3, in 16 cases, accidentally, is equal to $(0 \cdot 6)^{16} = 1/3500$.

---

* 'Phys. Rev.,' vol. 46, p. 633 (1934).

An atomic weight is equal to an integer N plus a small fraction. Since the reaction equations have energies which are multiples of q_m we may suppose that an atomic weight A is given by $A = N(1 + b) + sq_m$, where b is a small fraction the same for all elements and s is an integer which may be called the energy integer. In the reaction equations the terms $N(1 + b)$ cancel out, so that the reaction energy E is given by $E = \Sigma sq_m = nq_m$. The values of b and q_m depend on the atomic weight scale used, but N and s do not. If the scale used is such that $b = 0$ then $A = N + sq_m$.

TABLE II

Reaction	Energy	Energy/q_m	nq_m
(1)	0·00386	9·30	0·003735
(2)	0·02368	57·07	0·023655
(3)	0·00537	12·94	0·005395
(4)	0·01831	44·13	0·018260
(5)	0·0161*	38·80	0·016185
(6)	0·00051	1·23	0·000415
(7)	0·00211*	5·08	0·002075
(8)	0·00748*	18·02	0·007470
(9)	0·00477*	11·49	0·004565
(10)	0·0050	12·05	0·004980
(11)	0·00536*	12·92	0·005395
(12)	0·0017*	4·10	0·001660
(13)	0·0091	21·93	0·009130
(14)	0·0087*	20·96	0·008715
(15)	0·00212*	5·11	0·002075
(16)	0·00285	6·87	0·002905
(17)	0·00426	10·26	0·004150

The 17 reaction equations and the six relations which they satisfy are equivalent to 11 independent equations, so that two of the 13 values of s for the 13 elements involved may be arbitrarily selected. The atomic weight of He^4 or 4·0034 is nearly equal to $4 + 8q_m$ and the atomic weight of H^1 or 1·0080 is nearly equal to $1 + 19q_m$. We therefore take $s = 8$ for He^4 and $s = 19$ for H^1, so that on an atomic weight scale for which $b = 0$ the atomic weights of He^4 and H^1 should be 4·00332 and 1·00789.

TABLE III

n^1	20	He^3	41	Be^8	17	B^{10}	34
H^1	19	He^4	8	Be^9	33	B^{11}	27
H^2	34	Li^6	39	Be^{10}	36		
H^3	39	Li^7	41				

The reaction equations may then be solved for the other 11 values of the integer s. The values obtained in this way are given in Table III.

These integers, of course, give the correct values of n for all the 17 reactions. For example, reaction (6) or $Be^9 + H^1 \rightarrow Be^8 + H^2$ gives $33 + 19 = 17 + 34 + n$, which gives $n = 1$, and reaction (2) or $Li^6 + H^2 \rightarrow 2He^4$ gives $39 + 34 = 16 + n$, so that $n = 57$.

The atomic weight of helium with $O^{16} = 16$ may not be exactly equal to 4·00332. Bethe* got $He^4 = 4 \cdot 00336$ and Oliphant, Kempton, and Rutherford (*loc. cit.*) got $He^4 = 4 \cdot 0034$ with $O^{16} = 16$. With $He^4 = 4 + 8q_m$ the atomic weight of O^{16} should be $16 + s'q_m$, where s' is the energy integer for O^{16}. The atomic weight of He^4 with $O^{16} = 16$ is then $16\,(4 + 8q_m)/(16 + s'q_m)$, which is very approximately equal to $4 + q_m\,(8 - s'/4)$. The atomic weight of He^4 with $O^{16} = 16$ is therefore determined by s'. Table IV gives the values of He^4 for several values of s'.

TABLE IV

s'	Atomic weight of He^4
2	4·003113
1	4·003216
0	4·003320
−1	4·003424
−2	4·003528

Table V gives the atomic weights of the other elements with $He^4 = 4 \cdot 00332$ and 4·003424 and also the atomic weights given by Bethe (*loc. cit.*) and by Oliphant, Kempton, and Rutherford (*loc. cit.*). The fractional parts of the atomic weights with $He^4 = 4 \cdot 00332$ are equal to sq_m,

TABLE V—ATOMIC WEIGHTS WITH $O^{16} = 16$

Element	$He^4 = 4 \cdot 00332$	$He^4 = 4 \cdot 003424$	Bethe	Oliphant, Kempton, and Rutherford
n^1	1·00830	1·00833	1·0085	1·0083
H^1	1·00789	1·00791	1·00807	1·0081
H^2	2·01411	2·01416	2·01423	2·0142
H^3	3·01619	3·01626	3·01610	3·0161
He^3	3·01702	3·01709	3·01699	3·0172
He^4	4·00332	4·003424	4·00336	4·0034
Li^6	6·01619	6·01634	6·01614	6·0163
Li^7	7·01702	7·01720	7·01694	7·0170
Be^8	8·00706	8·00726	—	8·0071
Be^9	9·01370	9·01393	9·0135	9·0138
Be^{10}	10·01494	10·01520	—	10·0149
B^{10}	10·01411	10·01437	10·0146	10·0143
B^{11}	11·01121	11·01149	11·0111	11·0110

---

* 'Phys. Rev.,' vol. 47, p. 634 (1935).

and the atomic weights with $He^4 = 4{\cdot}003424$ were got from those with $He^4 = 4{\cdot}00332$ by multiplying by $4{\cdot}003424/4{\cdot}00332 = 1{\cdot}000026$.

The atomic weights with $He^4 = 4{\cdot}00332$ and $4{\cdot}003424$, of course, give the same energies for nuclear reactions not involving oxygen, so that the reaction energies in Table II do not help to decide between them. The first set agrees somewhat better with the atomic weights of Bethe and of Oliphant, Kempton, and Rutherford. Bethe's result that $He^4 = 4{\cdot}00336$ with $O^{16} = 16$ supports the value $He^4 = 4{\cdot}00332$ with $O^{16} = 16$. It seems probable, therefore, that $He^4 = 4{\cdot}00332$ with $O^{16} = 16$ is correct, but a final decision must be deferred until more exact measurements are available. If $He^4 = 4{\cdot}00332$ with $O^{16} = 16$ then the fractional parts of the atomic weights with $O^{16} = 16$ are equal to multiples of q_m.

The reaction $O^{16} + H^2 \rightarrow N^{14} + He^4$ discussed by Cockcroft is interesting in this connexion.* The atomic weight of N^{14} according to Bethe with $He^4 = 4{\cdot}00336$ is $14{\cdot}0076$. With $He^4 = 4{\cdot}00332$ this makes $N^{14} = 14{\cdot}00746$, so that the energy integer for N^{14} is $0{\cdot}00746/0{\cdot}000415 = 18$. The energy of this reaction is not known accurately, but according to Cockcroft it is about 3×10^6 electron volts, so that $8q_m$ is a possible value for it. If the energy is $8q_m$ then the energy integer s' of O^{16} is given by $s' + 34 = 18 + 8 + 8$, which gives $s' = 0$, and so is consistent with the atomic weight $4 + 8q_m$ for He^4 with $O^{16} = 16$.

SUMMARY

The energies of 17 nuclear reactions are discussed and revised values of some of the energies are adopted, which make the 17 reaction equations consistent. The energies are shown to be nearly equal to multiples of $q = 0{\cdot}000415$ in atomic weight units. The atomic weights of the 13 light elements involved are calculated from the reaction energies, and values are obtained nearly equal to the values of Bethe and of Oliphant, Kempton, and Rutherford.

---

* 'Rep. Conf. Int. Union Phys., Lond.,' p. 124 (1934).

On the Ground State of (H_2), the Molecular Ion $(H_2 +)$ and Wave Mechanics

By O. W. RICHARDSON, F.R.S., Yarrow Research Professor of the Royal Society

(*Received October* 2, 1935)

1—THE VALUES OF THE FUNDAMENTAL CONSTANTS

It* is not, I think, very generally known how accurate the check between the results of wave-mechanics calculations and the experimental data is, which is afforded by the fundamental constants of the H_2 molecule. In all probability this is partly because the improvements in the determinations of the various quantities have been made in an uncoordinated way by people largely interested in diverse questions and partly perhaps to some small inaccuracies in the published data. However all this may be, the present time seems opportune for a critical discussion of this question.

There are various ways in which such tests can be carried out, but the most exact is by means of the following energy relation:

Spectroscopic ionization potential of (H_2) + total negative energy of the molecular ion (H_2+) = total negative energy of (H_2) = total negative energy of 2(H) atoms + dissociation energy of (H_2). The total negative energy of (H_2+) is got from the wave mechanics by solving Schroedinger's equation for this structure and the total negative energy of an (H) atom is equal to R, the Rydberg constant. So far as anything pretending to accuracy is concerned, the spectroscopic *ionization* potential of H_2 is an experimental quantity, being got from the analysis of the H_2 spectrum. The spectroscopic value is, however, confirmed by the results of an electrical determination by Bleakney.† Until recently the only reliable determinations of the dissociation energy D of H_2 were also experimental. D is the least accurately determined of these four quantities. The best experimental values until about a year ago were probably $D = 4 \cdot 55 \pm 0 \cdot 16$ electron volts by Bichowsky and Copeland‡ (1928) and $D = 4 \cdot 34$, with about similar accuracy, by Witmer§ (1926).

---

* In this paper I assume the equivalence of the following numerical values:— Rh = 109677·8 wave numbers = 13·530 ± 0·005 electron volts, 1 electron volt = 8106·3 wave numbers.

† 'Phys. Rev.,' vol. 40, p. 496 (1932).

‡ 'J. Amer. Chem. Soc.,' vol. 50, p. 1315 (1928).

§ 'Phys. Rev.,' vol. 28, p. 1223 (1926).

The most recent value is that of Lozier,* $4 \cdot 43 \pm 0 \cdot 03$ e.v. This involves an experimental determination of the dissociation energy $D(H_2+)$ of the molecular ion (H_2+) and depends upon a measurement of the kinetic energy T_1 of the ions $(H+)$, formed when (H_2) is broken up by the impact of electrons of energy V_1 e.v. It also involves an experimental determination of the difference $V_1 - I(H_2)$, where $I(H_2)$ is the ionization potential of (H_2). $D(H_2+)$ is then got from the relation $D(H_2+) = V_1 - I(H_2) - 2T_1$. From this value of $D(H_2+)$, the value of D is got from the relation $D = I(H_2) - I(H) + D(H_2+)$. Lozier uses $I(H_2) = 15 \cdot 37$ e.v., the value Dr. Davidson and I found in 1929 (*see* below), to get $D = 4 \cdot 43 \pm 0 \cdot 03$ e.v. If the value $I(H_2) = 15 \cdot 334$ which we got by our third and later method (also considered below) were used this would reduce his value to $4 \cdot 39_4 \pm 0 \cdot 03$. In view of the various subtractions it seems not unlikely that this possible error of $\pm 0 \cdot 03$ e.v. is somewhat under-estimated. In any event Lozier's value is very close to the mean $4 \cdot 44_5 \pm$ *ca.* $0 \cdot 1$ of Bichowsky and Copeland's and Witmer's.

Dr. Davidson and I have published three determinations of D by subtracting 2R from the sum of the first two quantities in the first equation above but using slightly different methods of determining the spectroscopic ionization potential of H_2. The first two values were published† as $4 \cdot 476$ and $4 \cdot 450$ with a mean value of $4 \cdot 46$. However, the first value is affected by an arithmetical error and should have been $4 \cdot 45_6$ and the mean $4 \cdot 45_3$. This erroneous mean value $4 \cdot 46$ has frequently been quoted since both by ourselves and others. The third value,‡ which uses a method that we regard as theoretically superior to that of the first two, gave $4 \cdot 428$. I have gone over these calculations again in my book "Molecular Hydrogen and its Spectrum" (p. 153), using what I thought were rather better values of the data with these results: first method, $4 \cdot 45_7$; second method, $4 \cdot 45_4$; and third method, $4 \cdot 428$ e.v. Each of these was subject to a known possible error of $\pm 0 \cdot 04$ e.v. made up as follows: $\pm 0 \cdot 005$ from the spectroscopic ionization potential, $\pm 0 \cdot 03$ from Burrau's§ calculation of the negative energy of the molecular ion, and $\pm 0 \cdot 005$ from the value of 2R. However, since all three methods use Burrau's value and 2R in the same way the difference between the first two methods, on the one hand, and the third, on the other, cannot be due to chance errors in the experimental data on which the determinations are based, which should not exceed ± 0·005 e.v., but must be systematic and due to the different ways in which the data are treated.

---

* 'Phys. Rev.,' vol. 44, p. 575 (1933).

† 'Proc. Roy. Soc.,' A, vol. 123, p. 466 (1929).

‡ 'Proc. Roy. Soc.,' A, vol. 142, p. 63 (1933).

§ 'K. danske vidensk. Selsk.,' vol. 7, No. 14 (1927).

Light is thrown on this question by some recent developments. Hylleraas* has calculated the negative energy (W'_b) of the ion (H_2+) with fixed nuclei by a different method and changed it from 16·30 ± 0·03 to 1·20527 Rh = 16·3073 e.v. This has recently been checked by Jaffé,† using a third method and found correct to within about 0·0001 e.v. The interval between the ground state of (H_2) and the K = 1 $v = 0$ level of $1s\sigma\, 2p^1\, \Sigma$, computed on p. 156 of "Molecular Hydrogen and its Spectrum," is $\nu = 90233 \cdot 5$ wave numbers. The lowest rotational interval K (1) − K (0) of $v = 0 \cdot 2p^1\, \Sigma$ is 38·85 wave numbers. Thus the interval between the ground state of (H_2) and K = 0, $v = 0$ of $2p^1\, \Sigma$ is 90194·5 wave numbers. If we add to this the interval 32927 wave numbers between the W'_b of (H_2+) and K = 0, $v = 0$ of $2p^1\, \Sigma$ ("Molecular Hydrogen," etc., p. 161) we find for the interval between the ground state of (H_2) and the W'_b of (H_2+) 123121·5 wave numbers. Adding this to Hylleraas's W'_b = 132191·4 wave numbers, we find for the total negative energy of (H_2) in the ground state 255312·9 wave numbers. Subtracting 2Rh = 219355·6 wave numbers from this gives for the dissociation energy D (H_2) of (H_2) 35937·3 wave numbers = 4·436 e.v. The possible error in this value cannot be stated very definitely. It comes mainly from uncertainties in some of the finer points of the spectroscopic interpretation. It is probably conservative to say that this value of D (H_2) is reliable to ± 0·01 e.v.

This result can be checked by the following alternative method: subtract from Hylleraas's W'_b of (H_2+) 16·3073 ± 0·0001 e.v. the zero-point energy of (H_2+) 0·1399 ± 0·0007 e.v. (my estimate based on data in "Molecular Hydrogen and its Spectrum," p. 250). This gives the negative energy of the ground level of (H_2+) as 16·1674 ± 0·0008 e.v. Adding the spectroscopic ionization potential of (H_2), 15·334 ± 0·005 e.v., gives for the total negative energy of (H_2) in the ground state 31·5014 ± 0·006 e.v., and subtracting 2Rh = 27·060 ± 0·005 e.v. gives D (H_2) = 4·4414 ± 0·011, say 4·441 ± 0·011 e.v. The small discrepancy between this and the previous computation comes partly from the inaccuracy in the value of the zero-point energy of (H_2+), but this does not seem adequate to account for the whole of it.

The old first method (p. 504), using Hylleraas's W'_b of (H_2+) instead of Burrau's, gives for D (H_2) 4·4727 e.v., and the old second method with the same change 4·4705. The mean of these is 4·4716 e.v. This exceeds the value 4·4387 e.v. now got by the third method by $0\cdot032_9$ e.v.

---

* 'Z. Physik,' vol. 71, p. 739 (1931).

† 'Z. Physik,' vol. 87, p. 535 (1934).

James and Coolidge* have recently succeeded in calculating the binding energy of H_2 in the ground state directly on the wave mechanics. Their final result gives for the dissociation energy $D(H_2)$ the value 4·454 ± 0·013 e.v. With their estimate of accuracy this would set as upper and lower limits 4·467 and 4·441 e.v. respectively.

The value of James and Coolidge, 4·454, is identical with the mean $4\cdot455_1$ of our two values, viz., $4\cdot438_7$ got when the Rydberg (or similar) formula is supposed to apply when the nuclei are held at a fixed distance apart, and 4·4716 got on the assumption that it is the actual spectroscopic ν_0's which are governed by the Rydberg formula. As a confirmation of the principles of the wave mechanics this agreement is very satisfactory. But there are some points of view from which the situation is not so satisfactory.

I had expected when I first began to reconsider this matter that the recent improvements in the wave-mechanics calculations would settle which of the alternative assumptions about the Rydberg formula was correct. It seems, however, that this is just what they do not do. The most probable value 4·4387 for $D(H_2)$ calculated by the fixed nuclei method is just below the lower limit of James and Coolidge 4·441, whilst our earlier value 4·4716 got by the alternative set of assumptions is just above their upper limit. Either set of assumptions would give a set of possible values overlapping over a considerable range of the possible values resulting from the calculations of James and Coolidge, but there seems to be nothing as yet favouring one set rather than the other. Actually, it looks, so far, as though the Rydberg formula adopted a compromise between the two extreme sets of assumptions which we have considered. It is possible that a critical re-examination of the question would remove this anomaly by showing, for example, that something had been overlooked. On the other hand, it is possible that we have reached the limit of accuracy within which some of the arguments used in elucidating the data may be considered decisive. All the values of D that I have been discussing are assembled in Table I.

From Table I it is clear that Lozier's electrical value is identical with Richardson and Davidson's value got by method 3 when both sets of data are treated in a compatible way. The mean of all the most probable values given in the second column is 4·43976. It is, of course, not very logical to attach equal weight to all these data, particularly as four of them consist of pairs which must be mutually inconsistent. However, it is not absolutely certain which alternative should be chosen or whether the true value may not lie between in these cases; so that it is probable that the mean will lie nearer the true value than do the most extreme values in such a list. I admit that one cannot be certain on the point, but I am inclined to the view that the real value of D is very close to 4·440 e.v. in agreement with Richardson and Davidson's spectroscopic method 3 and with Lozier's electrical determination, and that James and Coolidge's wave-mechanics value is still a little too high by about the amount of the possible error 0·013 e.v. which they themselves estimate.

---

* 'J. Chem. Phys.,' vol. 1, p. 825 (1933); vol. 3, p. 129 (1935).

TABLE I—VALUES OF THE DISSOCIATION ENERGY OF (H_2) IN ELECTRON VOLTS

Method	Most probable value	Probable upper limit	Probable lower limit
James and Coolidge (wave mechanics)	$4\cdot45_4$	$4\cdot46_7$	$4\cdot44_1$
Richardson and Davidson (spectroscopic 1 and 2)	$4\cdot47_{16}$	$4\cdot48_{16}$	$4\cdot46_{16}$
Richardson and Davidson (spectroscopic 3)	$4\cdot43_{87}$	$4\cdot44_{87}$	$4\cdot42_{87}$
Bichowsky and Copeland (thermal)	$4\cdot5_6$	$4\cdot7_1$	$4\cdot3_9$
Witmer (spectroscopic, vibrational extrapolation)	$4\cdot3_4$	$4\cdot5_0$	$4\cdot1_8$
Lozier (electrical but using R. and D.'s I (H_2) by spectroscopic 1 and 2)	$4\cdot4_3$	$4\cdot4_6$	$4\cdot4_0$
Lozier (electrical but using R. and D.'s I (H_2) by spectroscopic 3).....................	$4\cdot3_{94}$	$4\cdot4_{24}$	$4\cdot3_{64}$

[*Note added in proof, October* 28, 1935—When the above was written I was unaware of two recent important determinations of D (H_2) by H. Beutler.* In fact, the second of them had not then been published. In the first paper he determines D (H_2) from the convergence of a new ultra-violet resonance band progression ($2p\sigma\,{}^1\Sigma_u{}^+$, $v' = 10$, $K' = 2 \rightarrow 1s\sigma\,{}^1\Sigma_g{}^+$, $v'' = 1$ to 14, $K'' = 1$ and 3). This is essentially the same method as that applied by Witmer to the Lyman progression. The advantage is that Beutler's progression has more vibrational states of $1s\sigma{}^1\Sigma_g{}^+$, the ground state of H_2, and so avoids Witmer's extrapolation. By this method Beutler finds D (H_2) $= 4\cdot454 \pm 0\cdot005$ e.v. ($= 36100 \pm 40$ cm^{-1} or $102\cdot68 \pm 0\cdot13$ kg cal).

In the second paper he reinvestigates the ultra-violet absorption spectrum of (H_2) at hydrogen pressures lower than those used by previous workers. As a result he finds that under the conditions of the examination by Dieke and Hopfield* the continuous absorption sets in at the rotational level K = 3 of the ground state of (H_2) and not at the K = 0 level as had been supposed. His redetermination of D (H_2) by Dieke and Hopfield's method raises the value from their 4·38 e.v. (accuracy not estimated) to $4 \cdot 445_5 \pm 0 \cdot 0008$ e.v. ($= 36116 \pm 6$ cm^{-1} wave numbers, or $102 \cdot 72 \pm 0 \cdot 02$ kg cal).

---

* 'Z. phys. Chem.,' vol. 27, p. 287 (1934); vol. 29, p. 315 (1935).

In this second paper he also discusses a determination of D (H_2) by Dieke and Blue.† In this paper they explain the absence of P, R branches observed by Richardson and Das in the system $3p^3\Pi \rightarrow 2s^3\Sigma$ of (H_2) at $v' \geqslant 4$ as due to predissociation. From a comparison of the vibrational levels at which these P, R branches disappear in (H_2), (HD), and (D_2) they were able to show that D (H_2) lies between 4·43 and 4·52 e.v. Beutler has refined this by pointing out that at the $v' = 3$ level of these (H_2) bands Richardson and Das found that whereas the lines involving K′ = 0, 1, 2, and 3 were sharp and of normal intensity the lines involving K′ = 4 (R3 and P5 lines) were diffuse and weak. From this he infers that predissociation occurs between the K′ = 3 and K′ = 4 levels at $v' = 3$ in (H_2) and thus narrows the limits found for D (H_2) by this method to 4·47 and 4·50 e.v.

My recalculation above of D (H_2), using Hylleraas's value of the W'_b of (H_2+) and assuming that the Rydberg formula applies to the W'_b's of the excited H_2 molecules with different principal quantum numbers when the nuclei are held at a fixed separation, certainly puts D (H_2) within the limits 4·425 and 4·452 e.v. and with some probability between 4·433 and 4·445 e.v. If Beutler's estimate of his accuracy is accepted the limits for his second and most accurate method are $4 \cdot 444_7$ and $4 \cdot 446_3$. Unless his estimates of accuracy are entirely wide of the mark, a supposition for which I can see no reasonable grounds, this makes the conclusion at which I had tentatively arrived from the consideration of the other data practically a certainty. These two sets of values are clearly identical within the limits of uncertainty involved in their respective determinations. This result vindicates the correctness of applying the Rydberg formula to the W'_b's with fixed nuclei rather than to the actual empirical ν_e's of the excited states. It is true that the predissociation method of Dieke and Blue leads to higher values which are more in accordance with those got by the first two methods of Richardson and Davidson; but this result does not appear to be supported by any other convincing evidence, and I feel fairly confident that the discrepancy will disappear with further investigation.]

---

* ' Phys. Rev.,' vol. 30, p. 400 (1927).

† ' Phys. Rev.,' vol. 47, p. 261 (1935).

2—THE MOLECULAR ION (H_2+)

The status of this interesting body, the simplest of all molecules, of which the ground state has the spectroscopic designation $1s\sigma\,{}^2\Sigma_g$, is most unusual; in fact, so far as I know, it is unique. Its actual existence is well established in the positive rays. Indeed, it is a familiar object to experimenters in this field and was one of the first of the new bodies to be discovered by Sir J. J. Thomson* in the early days of positive ray analysis. Except for the conclusion reached by Richardson and Tanaka†‡ from a study of low voltage electron discharges, that the reaction which occurs when H_2 is bombarded with electrons with energy in the neighbourhood of $I(H_2) = 15 \cdot 9$ volts is $(H_2) \rightarrow (H_2+) +$ electron, with the positive rays our direct knowledge of its actual existence begins and ends.§ None of its properties has ever been directly measured apart from such as are accessible to positive ray methods. It has no known discrete spectrum.

In spite of these drawbacks a great deal can be ascertained about its properties, particularly in the ground state. This information about the ground state can be obtained by two mutually independent methods. The first of these is an essentially experimental method. It depends on the fact that the ground state $1s\sigma\,{}^2\Sigma$ of the molecular ion is the limit to which a large number of singlet and triplet states of the neutral molecule (H_2), of the type $1s\sigma n\,(s, p, d, \ldots)(\sigma, \pi, \delta, \ldots)^{1\,3}\Sigma\,{}^{13}\Pi\,{}^{13}\Delta \ldots$, tend as the principal quantum number n tends to infinity. In any one of such series, as, for example, $1s\sigma np\pi\,{}^3\Pi$ ($n = 2, 3, 4, \ldots$) as n increases, the numbers which measure any given property of the states tend, and rather rapidly, to approach the value characteristic of the last state $1s\sigma \infty p\pi\,{}^3\Pi$, that is to say, the state $1s\sigma\,{}^2\Sigma$, the ground state of the ion.

The second method of obtaining such information is dependent on the comparatively very simple structure of the molecular ion which consists only of two protons and one electron. The Schroedinger equation for this structure can be completely solved and results of high numerical accuracy obtained without too much expenditure of time and trouble.

---

* 'Phil. Mag.,' vol. 13, p. 571 (1907).

† 'Proc. Roy. Soc.,' A, vol. 106, p. 640 (1924).

‡ A conclusion also reached independently by Smyth using a positive ray method (' Proc. Roy. Soc.,' A, vol. 105, p. 116 (1924); ' Phys. Rev.,' vol. 25, p. 452 (1925)).

§ Perhaps the determination of the dissociation energy of (H_2+) made by Lozier as indicated in § 1 should be considered as an exception to this statement but, even here the fact that one of the bodies on which observation is made is (H_2+) is an inference from other experiments.

As regards the ground state, to which I shall confine my attention for the present, the most accurate work is that of Burrau, Hylleraas, and Jaffé.

In the Silliman Lectures at Yale in 1932 I used the first (experimental) method to evaluate a number of the quantities which measure the more important properties of the ground state of the molecular ion. I also endeavoured to assess the limits within which these estimations might be considered reliable. In assessing these limits I did not employ any definite rules. They are just judgments about the consistency with which the values adopted as most probable seemed to be indicated by the various methods available. These conclusions have since been published in " Molecular Hydrogen and its Spectrum," most of them on p. 251 and in the final table.

I have collected these values and a few others which follow from them, either in an obvious way or in virtue of statements in the book referred to, in the second column of Table II below.

I shall now turn to the second method of evaluating such quantities, that is to say, the purely theoretical method which rests solely on the solution of Schroedinger's equation for the case of an electron in the field of two protons. At the time when the foregoing estimations were made there was no information by this method about a number of the quantities in Table II, although approximate values of several of them could be deduced from Burrau's computation of the negative energy of the ion in the ground state as a function of internuclear distance. The uncertainty in those theoretical estimations arose from the uncertainty, about $\pm 0{\cdot}03$ e.v., in Burrau's determination of the value $16{\cdot}29$ e.v., of the W'_b of $1s\sigma\,{}^2\Sigma$, and from the uncertainty of the accuracy within which his curves for the negative energy against internuclear distance could be differentiated. The first source of uncertainty has been removed by the work of Hylleraas and of Jaffé. Hylleraas has recalculated the value of W'_b for $1s\sigma\,{}^2\Sigma$ as $1{\cdot}20527$ Rh, and this has been confirmed to within $0{\cdot}00001$ Rh by Jaffé using an independent method. The second source of uncertainty has been removed by an investigation by Sandeman* which has just appeared.

Sandeman, following Jaffé's method for determining the energy as a function of internuclear distance r, redetermines the equilibrium internuclear distance r_e and then obtains the following expression for the variable part of the energy E_v as a function of $\xi = (r - r_e)/r_e$

$$E_v = 0{\cdot}41046 \text{ Rh } \xi^2 (1 - 1{\cdot}6978\xi + 2{\cdot}0933\xi^2 - 2{\cdot}1865\xi^3 + 1{\cdot}9775\xi^4 - 1{\cdot}4663\xi^5 + 0{\cdot}6528\xi^6 \ldots). \quad (1)$$

* 'Proc. Roy. Soc., Edin.,' vol. 55, p. 72 (1935).

Up to this point no physical assumptions are involved beyond those involved in writing down Schroedinger's equation for an electron in the field of two protons in the form in which it is used by Burrau, Hylleraas, and Jaffé.

The quantities in the last column above the solid line in Table II and other molecular constants are determined by the values of coefficients in an expansion such as equation (1). In general the number of the constants which can be determined and the accuracy with which they are determinable is limited solely by the number of known terms in the equation and, of course, by the accuracy within which the coefficients are known. The fundamental molecular constants of Table II are functions of the coefficients of equation (1) only and, in effect, may be regarded as these coefficients dressed up in a new notation.

The most convenient way of making this change over, for a diatomic molecule, is probably that due to Dunham* which is adopted by Sandeman.

The values of the molecular constants as actually given in Sandeman's paper are expressed in terms of the definitions of the new quantum mechanics whereas the experimentally derived data of Table II are in terms of the old quantum mechanics, so that the two sets of values are not immediately comparable. In order to institute a comparison I have reduced Sandeman's values to those of the corresponding quantities on the old quantum mechanics. This transcription has not been carried out to the limit of accuracy which should be possible with the data published by Sandeman. I have usually stopped when the introduction of further terms would only cause changes which are negligible compared with the uncertainty in the values determined from the experimental data.

An examination of Table II shows that the theoretical values are all very close to the experimental values, and except for ω_0 and $x\omega_0$ they lie within the range of limits within which I had assessed the experimental values as probably reliable. It seems to me that this agreement is a real triumph for the wave-mechanics or its equivalent the new quantum mechanics. Although in mere accuracy this test may not be so stringent as that considered in § 1 the agreement covers such a wealth of detail that the whole effect is most convincing. The two methods of deducing the quantities are completely independent. The theoretical method involves nothing beyond writing down Schroedinger's equation for an electron and two protons and carrying out its logical conclusions. It makes no appeal at any stage to any other empirical knowledge. On the other hand, the experimental values are deduced from a study of the properties of a large number of states, not of the molecular ion, but of an entirely different body, the neutral hydrogen molecule. This method is entirely empirical and does not make use of any fundamental theory.† Another curious feature of the situation which has already been mentioned is that no discrete spectrum involving any state of the molecular ion is known; consequently none of the quantities can at present be measured by experiments in which this body is itself directly concerned.

---

* 'Phys. Rev.,' vol. 41, pp. 713, 721 (1932).

TABLE II

	Experimental value	Theoretical value
B_0 (rotational band constant)	$29 \cdot 1 \pm 0 \cdot 1$ wave numbers	$29 \cdot 13$ wave numbers
r_0 (internuclear distance)	$1 \cdot 070 \pm 0 \cdot 004$ Å	$1 \cdot 069$ Å
B_1 (rotational band constant at $v = 1$)	$27 \cdot 7 \pm 0 \cdot 3$ wave numbers	$27 \cdot 51$ wave numbers
α (constant in $B_v = B_0 - \alpha v$)	$1 \cdot 4 \pm 0 \cdot 4$ wave numbers	$1 \cdot 617$ wave numbers
D_0 (coefficient of 2nd term of rotational energy of vibrationless state)	$0 \cdot 0198$ wave number	$0 \cdot 0199$ wave number
ω_0 (fundamental vibration frequency)	2235 ± 10 wave numbers	$2253 \cdot 3$ wave numbers
$x\omega_0$ (coefficient of v^2 in expression for energy)	62 ± 3 wave numbers	$67 \cdot 8$ wave numbers
$D(H_2+)$ (dissociation energy of $1s\sigma\,{}^2\Sigma$)*	$2 \cdot 61_6 \pm 0 \cdot 04$ e.v.	$2 \cdot 6362$ e.v.
W'_b (total negative energy with protons at rest at equilibrium distance)	—	$16 \cdot 3073$ e.v.

* Lozier's later experimental value (see § 1) is $2 \cdot 59 \pm 0 \cdot 03$ e.v.

It is perhaps desirable to say a word or two more about the values of ω_0 and $x\omega_0$ for which the theoretical values lie outside the limits predicted from the experimental data. Until the time of preparing the Silliman Lectures I had not devoted much consideration to the matter but had, for no very compelling reason, vaguely regarded a value $\omega_0 = 2210$ wave numbers got from Burrau's theoretical data as the best available value of this constant. However, I noticed that the two most reliable empirical estimations, the convergence of the $1s\sigma np\pi\,{}^3\Pi$ series at ω_0 between 2245 and 2250 and of the singlet states studied by Davidson and myself* at about $\omega_0 = 2280$, were both very much higher than 2210, and I decided that this number ought to be raised. I realize now that I did not lift it high enough. As regards the difference between $x\omega_0 = 67\cdot8$ and 62 ± 3 I do not think the empirical data are either sufficiently extensive or reliable to establish any real discrepancy.

---

† Some of these statements are perhaps a little too drastic. Either method of deduction does at some stage involve the general body of the theory of the spectra of diatomic molecules. The statements do, however, bring out the essential point I wish to make and I do not see how to put it more accurately without a long discussion.

The fact that the discrete spectrum of the molecular ion has not been found is also accounted for by the wave mechanics. Its ground state has about the same degree of stability as that of most of the excited states of (H_2). The dissociation energy of the ion is equal to Hylleraas's W'_b minus the vibrational zero point energy minus the Rydberg constant. This comes to $2\cdot6362$ e.v. The dissociation energies of about 40 excited states of (H_2) with a single excited electron all lie between $2\cdot37$ and $3\cdot42$ e.v., and most of them are close to $2\cdot65$ e.v. The curves connecting energy and internuclear distance of the deeper, and consequently more important, excited states of the molecular ion have been worked out on the wave mechanics by Teller.† Of these $2s\sigma\,{}^2\Sigma$, $2p\sigma\,{}^2\Sigma$, $3p\sigma\,{}^2\Sigma$, $3d\pi\,{}^2\Pi$ and $4f\sigma\,{}^2\Sigma$ possess no minimum, and are therefore unstable and incapable of existence. $3d\sigma\,{}^2\Sigma$ has a distinct minimum and should be fairly stable as its dissociation energy is about $1\cdot5$ e.v. It lies lower than any of the states except the ground state and the unstable $2p\sigma\,{}^2\Sigma$. It could go down to the unstable state $2p\sigma\,{}^2\Sigma$, giving rise to a continuous spectrum; but to emit a discrete spectrum it would have to go down to the ground state $1s\sigma\,{}^2\Sigma$, a transition which is forbidden by the $d \rightarrow s$ prohibition. The state $2p\pi\,{}^2\Pi$ has a very flat minimum estimated to correspond to a dissociation energy of about $0\cdot05$ e.v. This is so very small as to make the definite existence of this state dubious. It is at any rate clear both from what has been said and from the spacing of the energy levels in Teller's diagram that the wave mechanics predicts no discrete bands of the molecular ion in regions of the spectrum where they have been looked for up to the present. On the other hand, Brasefield‡ has observed a strong continuous spectrum in the electrodeless discharge in hydrogen at very low pressures which he attributes to transitions in which one or more of the unstable states of the molecular ion is involved, and this is quite possible. The existence of the ion in the positive rays implies that it must exist in at least one state possessing a fair degree of stability. If there is more than one such stable state the absence of a discrete spectrum

* 'Proc. Roy. Soc.,' A, vol. 123, p. 466 (1929).

† 'Z. Physik,' vol. 61, p. 458 (1930).

‡ 'Phys. Rev.,' vol. 33, p. 925 (1929).

(to the extent to which this can be regarded as established experimentally) requires that transitions between any pair of such stable states should be forbidden. Wave mechanics indicates that there is one pair of such stable states and that they belong to categories between which transitions are forbidden.

In concluding, I wish to acknowledge the help I have received in correspondence with Dr. H. M. James and Dr. P. M. Davidson in connexion with some of the points discussed in § 1 of this paper and with Dr. I. Sandeman with regard to § 2.

SUMMARY

This paper consists of two parts. The first part is a critical discussion of the agreement between the values of the fundamental constants, such as the dissociation energy, of the ground state of the hydrogen molecule (1) as determined by observations, and (2) as calculated by wave mechanics. The second part is a comparison of the properties of the molecular ion (H_2+) (1) as predicted empirically from a comparative study of various excited states of (H_2) and (2) as predicted by wave mechanics.

Exchange of Energy Between Diatomic Gas Molecules and a Solid Surface

By J. M. JACKSON, Ph.D., and A. HOWARTH, M.Sc., Department of Mathematics, University of Manchester

(*Communicated by D. R. Hartree, F.R.S.—Received December* 12, 1934—*Revised July* 31, 1935)

1—INTRODUCTION

When gas molecules are incident on a solid the exchange of energy may be measured by Knudsen's accommodation coefficient. If the gas molecules before collision have a mean energy corresponding to the temperature T_2 and the solid surface is at temperature T_1, then the gas molecules after collision will have a mean energy corresponding to a temperature T'_2, T'_2 being a function of T_1 and T_2.

The accommodation coefficient α is then defined as

$$\alpha = \mathrm{Lt}_{T_1 \to T_2} (T'_2 - T_2)/(T_1 - T_2).$$

Hitherto* it has only been possible to develop a theory for monatomic gases, the atoms of which possess translational energy alone. This paper is an attempt to include the effects which must arise from the rotation of diatomic molecules. The first step in the calculation of the accommodation coefficient is to obtain the probability $p\,(W_r;\ nm \mid n_1 m_1)$ that when a gas molecule with transitional energy W_r and rotational energy corresponding to the quantum state m collides with a solid atom in quantum state n, a transition will take place to the state specified by the quantum numbers n_1 and m_1, the balance of energy being taken up by the translational motion of the molecule. The accommodation coefficient α is then found by averaging $p\,(W_r;\ nm \mid n_1 m_1)$ over all accessible values of W_r, n, m, n_1, m_1, with the proper weighting factors.

In order to calculate $p\,(W_r;\ nm \mid n_1 m_1)$, we have used the Born-Dirac perturbation theory. It has only proved possible to carry out the calculations for a repulsive potential field of the type Ce^{-ar}. The attractive field between the gas molecule and the surface has been neglected, though for hydrogen and oxygen molecules it must be far from being small. The theory presented in this paper may therefore be regarded as an investigation of the effect of the repulsive field only.

---

* Jackson, 'Proc. Camb. Phil. Soc.,' vol. 28, p. 136 (1932); Jackson and Mott, 'Proc. Roy. Soc.' A, vol. 137, p. 703 (1932); Jackson and Howarth, 'Proc. Roy. Soc.,' A, vol. 142, p. 447 (1933).

The distribution of energy amongst the atoms of the solid is taken to be that of an assembly of oscillators all of the same frequency ν as in Einstein's specific heat theory. This is done for simplicity, as it is known* that averaging the accommodation coefficient over the normal modes of the solid does not make much difference in the results.

2—The Wave Equation

We take the gas molecule to be a symmetrical diatomic molecule, consisting of two atoms each of mass $m/2$. The motion of the molecule is assumed to be confined to a plane passing through the normal to the surface. Though this is admittedly a crude approximation, it is the simplest by which we are enabled to include the effects of rotation. We denote by x, the co-ordinate of the centre of mass of the molecule referred to, the mean position of the solid atom and by θ, the orientation of the molecule, θ being zero when the molecule is parallel to the surface.

An atom in the solid surface will be treated as a simple harmonic oscillator, of frequency ν with its centre at $y = 0$, where y is the displacement, normal to the surface of the solid, of the solid atom from its mean position. Then, assuming that the potential energy of either atom in the field of the solid surface is dependent only on its normal distance from the surface, and that each atom is in a repulsive exponential field $V = Ce^{-ar}$, the potential energy V will be of the form

$$V(x, y, \theta) = Ce^{-a(x-y-\frac{1}{2}l\sin\theta)} + Ce^{-a(x-y+\frac{1}{2}l\sin\theta)},$$

where l is the length of the molecule. Then

$$V = 2Ce^{-a(x-y)}\cosh(L\sin\theta),$$

where $L = al/2$. Now $a \sim 4 \times 10^8$ cm^{-1} and $l \sim 10^{-8}$ cm, the actual values of l for hydrogen and oxygen being 0·75 and 1·2 A respectively. Hence L is about 2.

The exact wave equation for the problem is

$$\left[-H + \lambda y^2 + 2Ce^{-a(x-y)}\cosh(L\sin\theta) + \frac{h}{2\pi i}\frac{\partial}{\partial t}\right]\Psi = 0, \qquad (1)$$

where

$$H = \frac{h^2}{8\pi^2}\left(\frac{1}{m}\frac{\partial^2}{\partial x^2} + \frac{1}{M}\frac{\partial^2}{\partial y^2} + \frac{1}{I}\frac{\partial^2}{\partial\theta^2}\right)$$

---

* Jackson, 'Proc. Camb. Phil. Soc.,' vol. 28, p. 136 (1932); Jackson and Mott, 'Proc. Roy. Soc.,' A, vol. 137, p. 703 (1932); Jackson and Howarth, 'Proc. Roy. Soc.,' A, vol. 142, p. 447 (1933).

and

$$\lambda = 2\pi^2 M\nu^2,$$

M being the mass of the solid atom and I the moment of inertia of the molecule. The wave equation can only be solved by perturbation theory. The zero-order approximation to the wave equation is chosen so that it is separable in the variables. Two alternatives present themselves, (i) we include the rotational potential energy as a term C cosh (L sin θ), (ii) we ignore the rotational potential energy entirely. These alternatives represent the two limiting cases of the motion.

If we consider the classical motion of a diatomic molecule with translational energy E_t and rotational energy E_r moving in the potential field $Ce^{-ax}\cosh(L\sin\theta)$, the point, x_r, at which the molecule ceases to be capable of performing complete rotations is given by

$$V\left(x_r, \frac{\pi}{2}\right) - V(x_r, 0) = Ce^{-ax_r}[\cosh L - 1] = E_r.$$

At some point, x_t, the translational motion of the molecule will be arrested. If $x_r > x_t$ the molecule ceases to rotate and commences to oscillate before its translational motion is arrested, whereas if $x_r < x_t$ the converse is true. Thus under certain circumstances a molecule may be describing oscillations about a line parallel to the surface when it is closest to the surface of the solid. This is the type of approximation considered in (i) above and that considered in (ii) is applicable when oscillation never occurs. According as the relevant values of E_r are small or large compared with those of E_t, the first or the second alternative will be nearer the truth in that region of the co-ordinate space which gives the major contribution to the matrix elements (where the interaction energy is large). At a large distance from the solid the interaction energy becomes exponentially small and it is therefore immaterial whether the first or the second type of approximation is more accurate for these values of x, the contribution to the matrix elements will be effectively zero in either case. The results have therefore been investigated, using both cases of zero-order approximation.

If we consider case (i), in which the gas molecule describes oscillations about $\theta = 0$, the potential energy V contains the factor cosh (L sin θ), which with sufficient accuracy for this approximation may be replaced by $(1 + \frac{1}{2}L^2\theta^2)$, for the two expressions are identical for small values of θ, whilst for values of θ near π/2 both are large if L is not too small. For such values of θ the wave functions corresponding to either potential will be small and practically indistinguishable from zero, whilst for small values of θ they will be identical. The appropriate zero-order wave equation for this case is

$$\left[-\mathrm{H} + \lambda y^2 + 2\mathrm{C}e^{-ax} + \mathrm{CL}^2\theta^2 + \frac{h}{2\pi i}\frac{\partial}{\partial t}\right]\Psi_0 = 0. \qquad (2)$$

For case (ii) in which the gas molecule executes complete rotations the zero-order wave equation is taken to be

$$\left[-\mathrm{H} + \lambda y^2 + 2\mathrm{C}e^{-ax} + \frac{h}{2\pi i}\frac{\partial}{\partial t}\right]\Psi_0 = 0, \qquad (3)$$

the dependence of the potential energy on θ having been ignored entirely. Both these equations are separable and soluble in terms of known functions.

The transition probability we require, $p\,(\mathrm{W}_r;\, nm \mid n_1m_1)$, *i.e.*, transition probability per collision for a transition of the solid atom from state n to state n_1, together with a transition of the rotational energy of the gas molecule from state m to state m_1, the balance of the energy being taken up by the translational motion of the gas molecule, is given by*

$$p\,(\mathrm{W}_r;\, nm \mid n_1m_1) = 4\pi^2\,[\mathrm{V}\,(nm \mid n_1m_1)]^2\,\lambda'\,(\mathrm{W}_{r'})/h\mathrm{N}\,(\mathrm{W}_r) \qquad (4)$$

where r and r' refer to the initial and final states of the translational energy, $\lambda'\,(\mathrm{W}_{r'})$ is the number of quantum states of translational energy per unit energy range and $\mathrm{N}\,(\mathrm{W}_r)$ is the number of gas molecules incident on the solid surface per unit time, per unit area. The matrix element $\mathrm{V}\,(nm \mid n_1m_1)$ is the integral

$$\int_{-\infty}^{\infty}\int_{-\pi}^{\pi}\int_{-\infty}^{\infty} dx\, d\theta\, dy\; \psi_r(x)\,\phi_m(\theta)\,\chi_n(y)\,\mathrm{V}\,\psi_{r'}(x)\,\phi_{m_1}(\theta)\,\chi_{n_1}(y)$$

in which $\psi_r(x)$, $\phi_m(\theta)$, and $\chi_n(y)$ are the wave functions of the translation and rotation of the gas molecule and of the oscillation of the solid atom respectively. The perturbation energy V is

$$2\mathrm{C}\,[e^{-a(x-y)}\cosh(\mathrm{L}\sin\theta) - e^{-ax} - \tfrac{1}{2}\mathrm{L}^2\theta^2]$$

or

$$2\mathrm{C}\,[e^{-a(x-y)}\cosh(\mathrm{L}\sin\theta) - e^{-ax}]$$

according as equation (2) or (3) is taken to be the zero-order wave equation. Since all the wave functions are orthogonal, and since in any transition resulting in an exchange of energy the oscillator state must change, the only contribution to $\mathrm{V}\,(nm \mid n_1m_1)$ comes from the term

---

* Jackson, 'Proc. Camb. Phil. Soc.,' vol. 28, p. 144 (1932).

$2Ce^{-a(x-y)}\cosh(\mathrm{L}\sin\theta)$ in the above expressions. Hence $\mathrm{V}(nm \mid n_1m_1)$ may be written

$$2\mathrm{C}\,\mathrm{X}(r \mid r')\,\mathrm{Y}(n \mid n_1)\,\mathrm{Z}(m \mid m_1),$$

where

$$\left.\begin{aligned} \mathrm{X}(r \mid r') &= \int_{-\infty}^{\infty} e^{-ax}\,\psi_r(x)\,\psi_{r'}(x)\,dx, \\ \mathrm{Y}(n \mid n_1) &= \int_{-\infty}^{\infty} e^{ay}\,\chi_n(y)\,\chi_{n_1}(y)\,dy, \\ \text{and}\quad \mathrm{Z}(m \mid m_1) &= \int_{-\pi}^{\pi} \cosh(\mathrm{L}\sin\theta)\,\phi_m(\theta)\,\phi_{m_1}(\theta)\,d\theta \end{aligned}\right\}. \tag{5}$$

The functions $\phi(\theta)$ depend on which choice of zero-order approximation has been made. As in this theory we are restricting our considerations to one quantum transitions only, the already known* values of the first two integrals may be quoted.

$$\mathrm{X}(r \mid r') = \frac{ah^2}{64\pi m\mathrm{C}}\,\frac{(\sinh\pi q'\,\sinh\pi q)^{\frac12}}{(\cosh\pi q' - \cosh\pi q)}\,(qq')^{\frac12}\,(q'^2 - q^2), \tag{6}$$

where $q = 4\pi mv_i/ah$ and $q' = 4\pi mv_n/ah$, v_i and v_n being the velocities of the gas molecule before and after collision.

$$[\mathrm{Y}(n \mid n \pm 1)]^2 = a^2h^2\,(n + \tfrac12 \pm \tfrac12)/8\pi^2\mathrm{M}h\nu. \tag{7}$$

3—THE ROTATIONAL MOTION OF THE GAS MOLECULE

Taking equation (2) as the zero-order wave equation, the wave equation for the rotational motion of the gas molecule is

$$\frac{d^2\phi_m}{d\theta^2} + \frac{8\pi^2\mathrm{I}}{h^2}\,[\mathrm{W}_m - \tfrac12\mathrm{CL}^2\theta^2]\,\phi_m = 0\ ;$$

substituting $\kappa^2 = 8\pi^2\mathrm{ICL}^2/h^2$ and $\xi = (2\kappa)^{\frac12}\,\theta$, this becomes

$$\frac{d^2\phi_m}{d\xi^2} + \left[\mathrm{W}_m\,\frac{4\pi^2\mathrm{I}}{h^2\kappa} - \tfrac14\xi^2\right]\phi_m = 0. \tag{8}$$

We can replace the boundary conditions which should be imposed on the solutions of this equation at $\theta = \pm\pi$ by the usual boundary conditions at $\xi = \pm\infty$, since the solutions of equation (8) are vanishingly small when $\theta = \pm\pi$. The solutions of (8) are then the Hermitian orthogonal functions

$$\phi_m(\xi) = \left(\frac{\kappa}{\pi}\right)^{\frac14}(2m\,!)^{-\frac12}\,e^{-\frac14\xi^2}\,\mathrm{H}_m(\xi),$$

---

* Jackson and Mott, 'Proc. Roy. Soc.,' A, vol. 137, p. 703 (1932), equations (25) and (22).

where $H_m(\xi)$ is the Hermitian polynomial of order m. The energy levels, W_m, are given by $W_m = h^2\kappa(2m+1)/4\pi^2 I$; $m = 0, 1, \ldots$.

In evaluating the matrix element $Z(m \mid m_1)$, it is therefore permissible to expand $\cosh(L \sin\theta)$ as a power series in θ and retain only the first two terms, since the wave functions ϕ_m and ϕ_{m_1} are only large when θ is small. Hence $Z(m \mid m_1)$ may be written

$$\tfrac{1}{2}\left(\frac{\kappa}{\pi}\right)^{\frac{1}{2}}(m\,!\;m_1\,!)^{-\frac{1}{2}}\int_{-\infty}^{\infty}(1 + L^2\xi^2/4\kappa)e^{-\frac{1}{2}\xi}\;H_m H_{m_1}\frac{d\xi}{(2\kappa)^{\frac{1}{2}}},$$

which gives

$$\left.\begin{aligned} Z(m \mid m_1) &= L^2[m(m-1)]^{\frac{1}{2}}\;/4\kappa, \text{ if } m_1 = m-2,\\ &= 1 + L^2(2m+1)/4\kappa, \text{ if } m_1 = m\\ &= L^2[(m+2)(m+1)]^{\frac{1}{2}}/4\kappa, \text{ if } m_1 = m+2\\ &= 0 \end{aligned}\right\} \qquad (9)$$

for all other values of m_1.

If, however, we take (3) to be the zero-order wave equation, the wave equation for the rotational motion of the gas molecule becomes

$$\frac{d^2\phi_m}{d\theta^2} + \frac{8\pi^2 I}{h^2}W_m\phi_m = 0$$

with normalized solutions $\dfrac{1}{(2\pi)^{\frac{1}{2}}}e^{\pm im\theta}$ and energy levels given by

$$W_m = h^2m^2/8\pi^2 I, \quad m = 0, 1, \ldots \quad .$$

We have then to calculate the value of $Z(m \mid m_1)$ using these wave functions. From equations (5), writing $q = (m - m_1)$,

$$\begin{aligned} Z(m \mid m_1) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{\pm iq\theta}\cosh(L\sin\theta)\,d\theta\\ &= \frac{1}{\pi}\int_{-\pi}^{\pi}\cos q\theta\cosh(L\sin\theta)\,d\theta, \end{aligned}$$

the imaginary part of the integral vanishing as the integrand is an odd function of θ. It is found that

$$\begin{aligned} Z(m \mid m_1) &= 0, &&\text{if } q \text{ is odd,}\\ &= (-)^{\frac{1}{2}q}I_q(L), &&\text{if } q \text{ is even,} \end{aligned}$$

$I_q(L)$ is a Bessel function of the first kind, of order q and imaginary argument. The result is easily deduced from Bessel's integral for $J_n(z)$.

4—THE TRANSITION PROBABILITY PER COLLISION

We can now write down expressions for the quantity $p\,(W_r;\, nm \mid n_1m_1)$, derived in equation (4). Knowing that* $N\,(W_r) = ahq/8\pi m$ and $\lambda'\,(W_{r'}) = 8\pi m/ah^2q'$ we have

$$p\,(W_r;\, nm \mid n_1m_1) = \frac{\pi^2}{16}\,\frac{(q'^2 - q^2)\sinh \pi q \sinh \pi q'}{(\cosh \pi q' - \cosh \pi q)^2}\,Y\,(n \mid n_1)^2\,Z\,(m \mid m_1)^2$$

and in particular

$$p\,(W_r\;;\, nm \mid n \pm 1, m_1)$$

$$= \frac{8\pi^4}{a^2h^2}\,\frac{m^2\,(n + \frac{1}{2} \pm \frac{1}{2})}{Mh\nu}\,\frac{(W_r - W_{r'})^2 \sinh \pi q \sinh \pi q'}{(\cosh \pi q' - \cosh \pi q)^2}\,Z\,(m \mid m_1)^2 \qquad (10)$$

where $q = \frac{4\pi}{ah}\,(2mW_r)^{\frac{1}{2}}$ and $q' = \frac{4\pi}{ah}\,(2mW_{r'})^{\frac{1}{2}}$, W_r and $W_{r'}$ being the translational energies of the gas molecule before and after collision.

5—THE CALCULATION OF THE THERMAL ACCOMMODATION COEFFICIENT

In order to determine the value of α, the thermal accommodation coefficient, we must calculate the average kinetic energy of the gas molecules after collision and from this determine the accommodation coefficient by the limiting process introduced by one of us.†

Let $N\,(W_r)$ be the flux of gas molecules, a_n the fraction of the solid atoms in the state n, and b_m the fraction of gas molecules with rotational energy in state m. Then the number of gas molecules with translational energy W_r which collide with an oscillator in state n and change it to state n_1, the rotational energy of the gas molecule changing from state m to state m_1 at the same time, is $N\,(W_r)\,a_n\,b_m\,p\,(W_r;\, nm \mid n_1m_1)$. The translational energy per unit energy range, per unit area and per unit time which this set of molecules transports is $N\,(W_r)W_{r'}\,a_n\,b_m\,p\,(W_r;\, nm \mid n_1m_1)$, where $W_{r'} = W_r + (n - n_1)\,h\nu + (\varepsilon_m - \varepsilon_{m_1})\,h\nu$ is the translational energy after collision, $\varepsilon_m h\nu$ being its rotational energy in state m. The total energy carried away by the translational motion of the gas molecules with initial translational energy W_r is then

$$\sum_{n}\sum_{n_1}\sum_{m}\sum_{m_1} N\,(W_r)\,[W_r\,a_n\,b_m\,p\,(W_r;\, nm \mid n_1m_1)$$

$$+ (n - n_1 + \varepsilon_m - \varepsilon_{m_1})h\nu a_n\,b_m\,p\,(W_r;\, n_1m_1 \mid nm)].$$

---

* Jackson, 'Proc. Camb. Phil. Soc.,' vol. 28, pp. 149, 145 (1932).

† Jackson, 'Proc. Camb. Phil. Soc.,' vol. 28, p. 152 (1932).

Then, averaging over all accessible values of the translational energy W_r, we define a temperature T'_2 in the usual manner

$$(T'_2)^{3/2} = (T_2)^{3/2}\left[1 + \mu_2^2 \sum_n \sum_{n_1} \sum_m \sum_{m_1} (n - n_1 + \varepsilon_m - \varepsilon_{m_1}) \int_0^\infty a_n b_m \, p\,(W_r ;\, nm \mid n_1 m_1)\, e^{-\mu_2 E} d\,E\right],$$

where $\mu = h\nu/kT$ and $E = W/h\nu$.

Let us now consider the six possible changes: (*a*) $nm \mid n_1m_1$, (*b*) $n_1m_1 \mid nm$, (*c*) $nm_1 \mid n_1m$, (*d*) $n_1m \mid nm_1$, (*e*) $nm \mid n_1m$, and (*f*) $n_1m \mid nm$. The changes represented by (*a*) and (*b*), (*c*) and (*d*), (*e*) and (*f*) are each the reverse of the other. Grouping these changes in pairs, we obtain

$$(T'_2)^{3/2} = (T_2)^{3/2}\,[1 + \mu_2^2 \sum_{\substack{n \\ n > n_1}} \sum_{n_1} \sum_{\substack{m \\ m > m_1}} \sum_{m_1} (A + B + C)],$$

where

$$A = (n - n_1 + \varepsilon_m - \varepsilon_{m_1}) \int_0^\infty \{a_n\, b_m\, p\,(nm \mid n_1m_1) - a_{n_1} b_{m_1}\, p\,(n_1m_1 \mid nm)\}\, e^{-\mu_2 E}\, dE \tag{11A}$$

$$B = (n - n_1) \int_0^\infty \{a_n b_m\, p\,(nm \mid n_1m) - a_{n_1} b_m\, p\,(n_1m \mid nm)\}\, e^{-\mu_2 E}\, dE \tag{11B}$$

$$C = (n - n_1 + \varepsilon_{m_1} - \varepsilon_m) \int_0^\infty \{a_n b_{m_1}\, p\,(nm_1 \mid n_1m) - a_{n_1} b_m\, p\,(n_1m \mid nm_1)\}\, e^{-\mu_2 E}\, dE. \tag{11C}$$

Now in (11A) we may put

$$\begin{aligned} p\,(E ;\, n_1m_1 \mid nm) &= p\,(E - E') ;\, (nm \mid n_1m_1) \quad &\text{if } E > E' \\ &= 0 &\text{if } E < E', \end{aligned}$$

where $E' = (n - n_1 + \varepsilon_m - \varepsilon_{m_1})$. Similar substitutions may be made in (11B) and (11C), the collisions implied by the right-hand sides of these equations being the exact reversals of the collisions implied by the left-hand sides. Thus we may write

$$A = (n - n_1 + \varepsilon_m - \varepsilon_{m_1})\,\{a_n b_m - a_{n_1} b_{m_1}\, e^{-(n-n_1)\mu_2 - (\varepsilon_m - \varepsilon_{m_1})\mu_2}\} \int_0^\infty e^{-\mu_2 E}\, p\,(nm \mid n_1m_1)\, dE$$

$$B = (n - n_1)\,[a_n b_m - a_{n_1} b_m\, e^{-(n-n_1)\mu_2}] \int_0^\infty e^{-\mu_2 E}\, p\,(nm \mid n_1m)\, dE$$

$$C = [\,(n - n_1) - (\varepsilon_m - \varepsilon_{m_1})]\,[a_n b_{m_1} - a_{n_1} b_m\, e^{-\mu_2[(n-n_1)-(\varepsilon_m-\varepsilon_{m_1})]}] \int_0^\infty e^{-\mu_2 E}\, p\,(nm_1 \mid n_1m)\, dE.$$

The accommodation coefficient α is then defined by the equation

$$\alpha = \mathrm{Lt}_{\mathrm{T}_1 \to \mathrm{T}_2 \to \mathrm{T}}\ \tfrac{2}{3}\, \mathrm{T}_2 \mu_2^2 \sum_{\substack{n \\ n > n_1}} \sum_{n_1} \sum_{\substack{m \\ m > m_1}} \sum_{m_1} (\mathrm{A} + \mathrm{B} + \mathrm{C}) \Big/ \frac{h\nu}{k}\left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right). \tag{12}$$

The values to be assigned to b_m, the weight of the rotational state m in this expression, depend on the form of the wave function for the rotational motion of the gas molecule. The weights of the oscillator states are given by $a_n = e^{-n\mu}(1 - e^{-\mu})$. If we consider the case in which the rotator is effectively replaced by an oscillator, we must write for b_m, $b_m = e^{-m\lambda}(1 - e^{-\lambda})$ where $\lambda = h^2\kappa/4\pi^2 \mathrm{I}k\mathrm{T}$.

Considering one quantum transitions only, and writing $\alpha(1)$ for the partial accommodation coefficient corresponding to a change of quantum number of magnitude unity in the oscillator states of the solid, we find, on substituting for a_n and b_m in (12) and proceeding to the limit, that

$$\begin{aligned}\alpha(1) = \tfrac{2}{3}\mu^3(e^{\mu} - 1)^{-1}(1 - e^{-\lambda})\frac{32\pi^4 m^2 h\nu}{a^2 h^2 \mathrm{M}} \sum_m e^{-m\lambda}\big[(1+2\delta)^3 \mathrm{Z}(m|m-2)^2 \mathrm{P}(\mu, +1) \\ + \mathrm{Z}(m|m)^2 \mathrm{P}(\mu, 0) + e^{\mu - 2\lambda}(1 - 2\delta)^3 \mathrm{Z}(m|m+2)^2 \mathrm{P}(\mu, -1)\big]\end{aligned} \tag{13}$$

where μ is the limiting value of μ_1 and μ_2 as $\mathrm{T}_1 \to \mathrm{T}_2 \to \mathrm{T}$ and

$$\delta = h^2\kappa/4\pi^2 \mathrm{I} h\nu, \quad \lambda = \mu\delta$$

and

$$\mathrm{P}(\mu, l) = \int_0^{\infty} \frac{\sinh \frac{\gamma}{a}(\mathrm{E}_l)^{\frac{1}{2}} \sinh \frac{\gamma}{a}(\mathrm{E})^{\frac{1}{2}}}{\left(\cosh \frac{\gamma}{a}(\mathrm{E}_l)^{\frac{1}{2}} - \cosh \frac{\gamma}{a}(\mathrm{E})^{\frac{1}{2}}\right)^2} e^{-\mu \mathrm{E}}\, d\mathrm{E},$$

where l is $+1$, 0, or -1 in the three integrals above, and E_{+1}, E_0, E_{-1}, and γ are defined by the equations

$$\mathrm{E}_{+1} = \mathrm{E} + 1 + 2\delta\,; \quad \mathrm{E}_0 = \mathrm{E} + 1\,; \quad \mathrm{E}_{-1} = \mathrm{E} - 1 + 2\delta,$$
$$\gamma^2 = 32\pi^4 m h\nu/h^2.$$

Substituting for Z in (13) from (9) and summing over all values of m, we obtain

$$\begin{aligned}\alpha(1) = \tfrac{2}{3}\frac{m}{\mathrm{M}}\mu^3(e^{\mu} - 1)^{-1}\frac{\gamma^2}{a^2}\Big[\frac{\mathrm{L}^4}{8\kappa^2}(1 + 2\delta)^3(e^{\lambda} - 1)^{-2}\mathrm{P}(\mu, +1) \\ + \left\{\left(1 + \frac{\mathrm{L}^2}{2\kappa}\right) + \mathrm{L}^2(e^{\lambda} - 1)^{-1} + \frac{\mathrm{L}^4}{16}(1 + 6e^{\lambda} + e^{2\lambda})(e^{\lambda} - 1)^{-2}\mathrm{P}(\mu, 0)\right\} \\ + \frac{\mathrm{L}^4}{8\kappa^2}[2\delta - 1]^3 e^{\mu}(e^{\lambda} - 1)^{-2}\mathrm{P}(\mu, -1)\Big].\end{aligned} \tag{14}$$

If we now consider the case in which the gas molecule behaves as a plane rotator, we must write for b_m, $b_m = e^{-m^2\lambda_1}/\mathrm{R}$, where $\mathrm{R} = 1 + 2\sum_{m=1}^{\infty} e^{-m^2\lambda_1}$ and $\lambda_1 = h^2/8\pi^2\mathrm{I}k\mathrm{T}$. We then find

$$\alpha(1) = \tfrac{2}{3}\,\mu^3\,(e^{\mu} - 1)^{-1}\,\mathrm{R}^{-1}\,\frac{m}{\mathrm{M}}\,\frac{\Upsilon^2}{a^2}\sum_{m}\sum_{\substack{m_1\\ m > m_1}} [\mathrm{A}_1 + \mathrm{B}_1 + \mathrm{C}_1], \tag{15}$$

where

$$\mathrm{A}_1 = e^{-m^2\lambda_1}\,[1 + (m^2 - m_1^2)\,\beta]^3\,\mathrm{Z}\,(m \mid m_1)^2\,\mathrm{P}\,(\mu, +1)$$

$$\mathrm{B}_1 = e^{-m^2\lambda_1}\,[\mathrm{Z}\,(m \mid m_1)]^2\,\mathrm{P}\,(\mu\,|, 0)$$

$$\mathrm{C}_1 = e^{-m_1^2\lambda_1}\,[1 - (m^2 - m_1^2)\,\beta]^3\,\mathrm{Z}\,(m \mid m_1)^2\,\mathrm{P}\,(\mu, -1)$$

or

$$\mathrm{C}_1 = e^{\mu - m^2\lambda_1}\,[(m^2 - m_1^2)\,\beta - 1]^3\,\mathrm{Z}\,(m \mid m_1)^2\,\mathrm{P}\,(\mu, -1)$$

according as $1 - (m^2 - m_1^2)\,\beta \gtrless 0$. $\mathrm{P}\,(\mu, l)$ is defined as before, E_{+1}, E_0, and E_{-1} being given in this case by

$$\mathrm{E}_{+1} = \mathrm{E} + 1 + (m^2 - m_1^2)\,\beta; \quad \mathrm{E}_0 = \mathrm{E} + 1;$$

$$\mathrm{E}_{-1} = \mathrm{E} + |\,1 - (m^2 - m_1^2)\beta\,|,$$

where

$$\beta = \lambda_1/h\nu.$$

6—NUMERICAL RESULTS

The numerical value of the accommodation coefficient has been calculated for hydrogen and oxygen incident on a tungsten surface for the following values of the temperature, T,

410, 273, 205, 164, 102·5, 51·25° K.

The most probable value of the exponential index a has been shown* to be 4×10^8 cm^{-1}. The value of C we take to be 10^{-13} ergs. The characteristic temperature of the solid (tungsten) is 205° K, corresponding to an atomic frequency of $4 \cdot 3 \times 10^{12}$ sec^{-1}. This frequency is determined from Lindemann's melting-point formula.

Considering firstly the gas molecule as an oscillator, we may write

$$\alpha = \alpha(1) + \alpha(2) + \ldots,$$

where

$$\alpha(1) = p\mathrm{P}\,(\mu, +1) + q\mathrm{P}\,(\mu, 0) + r\mathrm{P}\,(\mu, -1). \tag{16}$$

In this expression p, q, and r are functions of T and are defined for both gases by equation (14). It is interesting to note at this stage that the term $q\mathrm{P}(\mu, 0)$ corresponds to transitions in which no change occurs in the rotational state of the gas molecule. It would appear at first sight that this term should give the same value for the accommodation coefficient as the theory for a monatomic gas. This, however, is not so, as this partial accommodation coefficient contains a factor derived from averaging the rotational potential energy over all directions. The difference between this result and the simple theory may therefore be ascribed to the finite dimensions of the molecule. This difference persists whichever zero-order approximation is used. The term $p\mathrm{P}(\mu, +1)$ corresponds to transitions in which the oscillator and the rotator both lose or gain energy, whilst the term $r\mathrm{P}(\mu, -1)$ corresponds to transitions in which the rotator gains energy and the oscillator loses energy or *vice versa*. The values of p, q, and r have been calculated for hydrogen and oxygen and are given in Table I.

---

* Jackson and Howarth, 'Proc. Roy. Soc.,' A, vol. 142, p. 447 (1933).

TABLE I

	Hydrogen			Oxygen		
T	p	q	r	p	q	r
410	0·04098	0·08260	0·02954	0·9926	19·85	0·03796
273	0·01042	0·1523	0·00936	0·7062	35·85	0·03469
205	0·00226	0·2294	0·00268	0·4734	52·87	0·02986
164	0·00050	0·3064	0·00067	0·3117	69·46	0·02525
103	—	0·4875	—	0·0896	107·2	0·01536
52	—	0·4649	—	0·0015	99·9	0·00184

Now consider the value of $\mathrm{P}(\mu, l)$ for various values of l.

$$\mathrm{P}(\mu, l) = \int_0^\infty \frac{\sinh \frac{\gamma}{a}(\mathrm{E}_l)^{\frac{1}{2}} \sinh \frac{\gamma}{a}(\mathrm{E})^{\frac{1}{2}}}{\left(\cosh \frac{\gamma}{a}(\mathrm{E}_l)^{\frac{1}{2}} - \cosh \frac{\gamma}{a}(\mathrm{E}_l)^{\frac{1}{2}}\right)^2} e^{-\mu\mathrm{E}} d\mathrm{E}.$$

We require some knowledge of the variation of

$$\frac{\sinh \frac{\gamma}{a}(\mathrm{E}_l)^{\frac{1}{2}} \sinh \frac{\gamma}{a}(\mathrm{E})^{\frac{1}{2}}}{\left(\cosh \frac{\gamma}{a}(\mathrm{E}_l)^{\frac{1}{2}} - \cosh \frac{\gamma}{a}(\mathrm{E})^{\frac{1}{2}}\right)^2}$$

with l. Now for hydrogen $\gamma/a = 6\cdot5$ and for oxygen, $= 26\cdot0$. Thus for values of E greater than (say) $\mathrm{E} \sim 0\cdot6$ we can write the above factor as

$$e^{-\frac{\gamma}{a}((\mathrm{E}_l)^{\frac{1}{2}} - (\mathrm{E})^{\frac{1}{2}})} \left(1 + 2e^{-\frac{\gamma}{a}((\mathrm{E}_l)^{\frac{1}{2}} - (\mathrm{E})^{\frac{1}{2}})} + \ldots\right).$$

From this expansion it is seen that the above function is monotonic and decreasing for fixed E with increasing l. Hence $P(\mu, l)$ also behaves in the same way. Similarly, for fixed l, $P(\mu, l)$ decreases as μ increases and more rapidly when l is large than when it is small. The magnitudes of the $P(\mu, l)$'s which enter into equations (14) and (16) depend on the relative magnitudes of E_1. It is found that for hydrogen

$$P(\mu, +1) < P(\mu, -1) < P(\mu, 0)$$

and for oxygen

$$P(\mu, +1) < P(\mu, 0) < P(\mu, -1).$$

Then, bearing in mind the relative magnitudes of the coefficients p, q, and r in the two cases, a fairly good approximation to $\alpha(1)$ will be given by

$$\alpha(1) = q\,P(\mu, 0). \tag{16A}$$

The effect of the rotation will be practically negligible compared with the above both on account of the small coefficient and the fact that the corresponding P's are smaller than $P(\mu, 0)$. The variation of $\alpha(1)$ with temperature given by equation (16) is shown for hydrogen and oxygen in fig. 1, curves 1 and 2. The conclusion to be drawn from these results is that where it is possible to treat the rotation of the gas molecule as an oscillator, no appreciable contribution to the exchange of energy comes from changes in the rotator state, except perhaps at high temperatures.

When the gas molecule is treated as a rotator, the integrals $P(\mu, l)$ depend on the values of m and m_1, and so no summation in finite terms is possible. Writing $\alpha(1 \mid n)$ for the partial accommodation coefficient corresponding to a change of magnitude unity in the quantum number of the oscillator together with a change of magnitude n in the quantum number of the rotator, we have from equation (15)

$$\alpha(1 \mid n) = \frac{2}{3}\,\mu^3 R^{-1} (e^{\mu} - 1)^{-1} \frac{m\gamma^2}{Ma^2} \sum_{m} \sum_{\substack{m_1 \\ m > m_1}} [A_1 + B_1 + C_1].$$

Considering oxygen first, we find that $\beta = 1{\cdot}04 \times 10^{-2}$. Thus $(m^2 - m_1^2)\,\beta \sim 0$, since when $(m^2 - m_1^2)$ is large either $[Z(m \mid m_1)^2]$ or $e^{-m^2\lambda_1}$ will be small. Hence for oxygen

$$\left.\begin{aligned} \alpha(1 \mid 0) &= \frac{2}{3}\,\mu^3 \frac{m\gamma^2}{Ma^2} R^{-1} (e^{\mu} - 1)^{-1} \sum_{m} e^{-m^2\lambda_1} [I_0(L)]^2 P(\mu, 0) \\ &\text{and} \\ \alpha(1 \mid 2) &= \frac{2}{3}\,\mu^3 \frac{m\gamma^2}{Ma^2} R^{-1} (e^{\mu} - 1)^{-1} \sum_{m} e^{-m^2\lambda_1} [I_2(L)]^2 P(\mu, 0) \end{aligned}\right\}. \tag{17}$$

In these equations λ_1 is small and the sums $\Sigma e^{-m^2\lambda_1}$ may be obtained by the Euler-Maclaurin formula.*

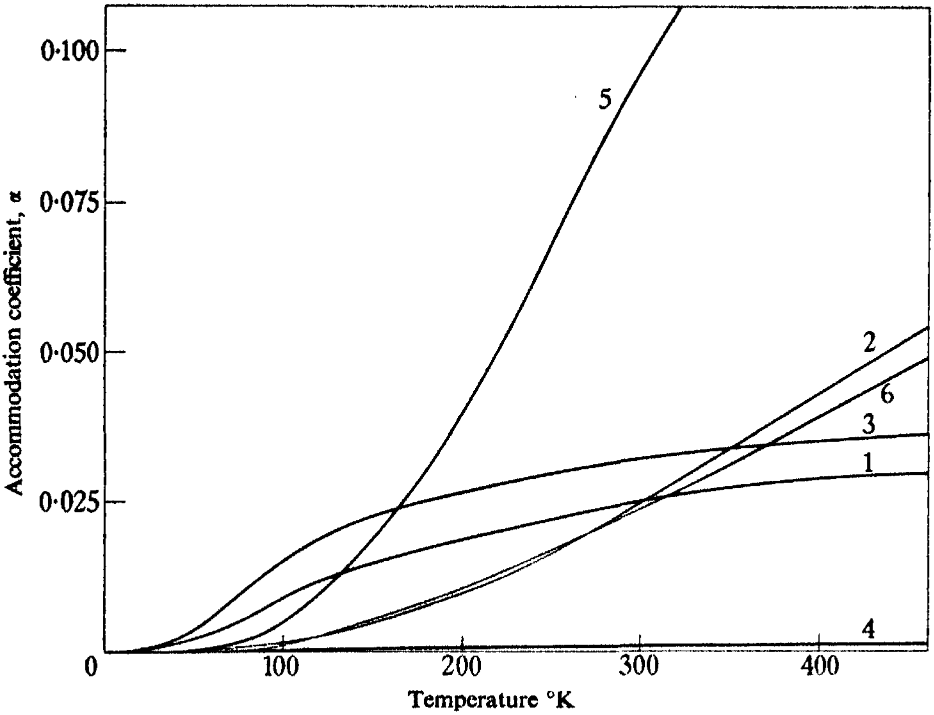


FIG. 1—Variation of the thermal accommodation coefficient α with temperature for hydrogen and tungsten, and for oxygen and tungsten. Curve 1, hydrogen, treating the gas molecules as oscillators. Curve 2, ditto for oxygen. For these two curves the effects of rotation are negligible. In curves 3–6 the gas molecules are treated as rotators. Curve 3, hydrogen, partial accommodation coefficient α (1/0). Curve 4, hydrogen, ditto α (1/2). Curve 5, oxygen, ditto α (1/0). Curve 6, oxygen, ditto α (1/2).

For hydrogen $\beta = 0{\cdot}426$, and we find $\alpha\,(1 \mid 0)$ is given by the same expression as for oxygen and

$$\alpha\,(1 \mid 2) = \frac{2}{3}\,\mu^3 \frac{m\gamma^2}{Ma^2}\,R^{-1}\,(e^{\mu} - 1)^{-1} \sum_{m=0}^{\infty} e^{-m^2\lambda_1}\,[\{1 + 4\,(m + 1)\,\beta\}^3\,P(\mu, +1)$$
$$+\, e^{\mu}\,\{4\,(m + 1)\,\beta - 1\}^3\,P\,(\mu, -1)]\,I_2^2\,(L).$$

Here the values of $e^{-m_1^2\lambda_1}$ will converge very quickly to zero, and whilst

---

* Whittaker and Robinson, " Calculus of Observations," p. 134. 2nd Ed. Blackie (1929).

$\{1 + 4(m+1)\beta\}$ will increase with m, $P(\mu, +1)$ will decrease. Thus we may write approximately

$$\alpha(1 \mid 2) = \frac{2}{3}\,\mu^3 \frac{m\gamma^2}{Ma^2} R^{-1} (e^{\mu} - 1)^{-1} [(1 + 4\beta)^3 P(\mu, 1 + 4\beta)\, e^{-4\lambda_1} + e^{\mu} (4\beta - 1)^3 P(\mu, 4\beta - 1)\, e^{-4\lambda_1} + e^{\mu} (8\beta - 1)^3 P(\mu, 8\beta - 1)\, e^{-9\lambda_1} + (1 + 8\beta)^3 P(\mu, 1 + 8\beta)\, e^{-9\lambda_1} + \ldots] \{I_2(L)\}^2. \tag{18}$$

The results have been calculated for hydrogen and oxygen from these formulae and are shown in fig. 1, curves 3–6. For hydrogen the effect of rotation $\alpha(1 \mid 2)$ is very small compared with $\alpha(1 \mid 0)$, for which there is no rotation. For oxygen which has a larger nuclear separation (1·2 A) and a larger moment of inertia the effects of rotation are more pronounced. The relatively large separation between the states of rotational energy of hydrogen inhibit the exchange of energy between the rotations of the gas molecule and the oscillations of the solid atoms. The exchange of rotational energy will therefore become more pronounced the larger the moment of inertia of the molecule becomes.

The type of approximation for which we have obtained this result is the well-known Born approximation in the theory of collisions. This always gives too large a result. We may therefore regard the two approximations considered above as the lower and upper bounds to the true result.

The results here obtained must only be regarded as a preliminary investigation of the problem of energy exchange between diatomic molecules and a solid surface. No account has been taken of the van der Waals attractive field of the solid. There is no satisfactory way of including this in a Born-Dirac perturbation theory, unless it is small, which will not be so for hydrogen and oxygen. A more refined treatment of collision problems due to London* will have to be used, in which it is possible to include large perturbations. The investigation of the problem by this method is now in progress.

In conclusion we wish to thank Professor Hartree for his continued interest in our work and for much helpful criticism during its progress.

SUMMARY

The theory of the accommodation coefficient given by Jackson and Mott has been extended for a gas composed of diatomic molecules. Two

---

* ' Z. Physik,' vol. 74, p. 143 (1932).

limiting cases have been discussed: (*a*) the gas molecule is treated as an oscillator, (*b*) as a plane rotator. In the first case the exchange of energy between the rotations of the gas molecule and the vibrations of the solid is negligible, whilst in the second the effect of the rotation is very small for hydrogen and larger for oxygen, taking the same repulsive field of the solid in the two cases. Under these conditions the energy exchange will be larger the larger the moment of inertia of the gas molecule.

Statistical Treatment of Strong Electrolytes

By S. LEVINE,* Department of Physics, University of Cincinnati

(*Communicated by J. C. McLennan, F.R.S.—Received March* 6—*Revised June* 14, 1935)

1—INTRODUCTION

By assuming the Coulomb forces between the ions of an electrolytic solution and by combining Poisson's equation and Boltzmann's principle, Debye and Hückel† have developed a theory of strong electrolytes, satisfactory, however, only up to moderate concentrations. The approximate character of their method has been extensively investigated by a number of authors. Proceeding from the laws of statistics, Fowler‡ showed that only for dilute solutions is Poisson's equation valid in the Debye-Hückel treatment, where phenomena on a molecular scale are involved, the omitted fluctuation terms becoming appreciable at stronger concentrations. Although Halpern§ claims that at no concentration of physical interest is such a neglect permissible, we may assume from more recent work|| that Fowler's original conclusions hold.

Another objection is the use of the Coulomb law of force between two ions of charge ε_i and ε_j,

$$F_{ij} = \varepsilon_i \varepsilon_j / D r_{ij}^2, \tag{1.1}$$

---

* Formerly at the McLennan Physical Laboratory, University of Toronto.

† 'Phys. Z.,' vol. 24, p. 185 (1923).

‡ "Statistical Mechanics," chaps. 8 and 13.

§ 'J. Chem. Phys.,' vol. 2, p. 85 (1934).

|| Kirkwood, *ibid.*, vol. 2, p. 767 (1934), and Fuoss, *ibid.*, vol. 2, p. 818 (1934). *See also* Onsager, 'Chem. Rev.,' vol. 13, p. 73 (1933).

which is not justifiable for small values of r_{ij}. In water, for example, the main part of D is contributed by the orientation of the water dipoles, and in the vicinity of an ion, a saturation effect will exist changing the value of D considerably. In addition, the polarization, van der Waals and exchange forces between the ions come into play at their close approach.*

To avoid the use of Poisson's equation, Kramers† has applied the methods of statistical mechanics evaluating directly Gibbs's phase-integral for the free energy F of the solution. Owing to the use of the approximate law (1.1), however, his expression for F is satisfactory only for very dilute solutions. In this article we shall be concerned with a more detailed analysis of Kramers's method.

Various authors‡ have extended the original Debye-Hückel theory to stronger concentrations. Hückel§ modified (1.1) by writing the dielectric constant as

$$D = D_0 - \sum_i \delta_i \gamma_i, \tag{1.2}$$

where γ_i is the concentration of ions of the ith kind and the δ_i's are empirical constants, so chosen that the activity coefficient is given correctly up to a concentration of a few moles per litre. However, there exist several objections|| to Hückel's treatment, and his results are useful only as interpolation formulae. Scatchard|| has shown in a qualitative manner that the form of (1.2) is doubtful, except for very dilute solution. He also indicated the importance of the non-ionic forces (*i.e.*, polarization and van der Waals forces) between the ions and water molecules.

Sack¶ worked out the change in "macroscopic" dielectric constant due to the saturation effect for very small concentrations, obtaining an expression similar to (1.2) but with much larger values for δ_i. Owing to the doubtful assumptions and the neglect of several factors, his results are to be questioned. Indeed the law** $D = D_0 + c\sqrt{\gamma}$ seems to be the correct one for dilute solutions. However, as the "macroscopic" dielectric constants may be quite different from the "microscopic"

---

* Any chemical hydration would also play an important part.

† 'Proc. Akad. Sci. Amst.,' vol. 30, p. 145 (1927).

‡ *See* Falkenhagen, "Elektrolyte," chap. 11.

§ 'Phys. Z.,' vol. 26, p. 93 (1925).

|| *See* Orthmann, "Ergebn. exakt. Naturwiss.," vol. 6, p. 155 (1927), and Scatchard. 'Phys. Z.,' vol. 33, p. 22 (1933).

¶ 'Phys. Z.,' vol. 28, p. 199 (1927). *See also* Debye, "Polar Molecules," chap. 6, and Orthmann, *loc. cit.*

** Debye and Falkenhagen, 'Phys. Z.,' vol. 29, p. 121, 401 (1928); Wien, 'Ann. Physik,' vol. 11, p. 429 (1931).

quantity involved in (1.2), this is not necessarily at variance with the form (1.2) used by Hückel.

More recently, Bonino* took account of the fluctuation terms in Poisson's equation by replacing the dielectric constant D_0 by the function

$$\Xi = D_0\left(1 - \frac{\sum_i \lambda_i x_i^3}{D_0}\right). \tag{1.3}$$

Here the λ_i's are constants and $x_i = \kappa a_i$, where a_i is the radius of the ion of the *i*th kind and κ^{-1} is the "characteristic length," in the Debye-Hückel theory. With given values of a_i and by suitable choice of λ_i, good agreement with experiment was obtained up to strong concentrations. The treatment, however, is of an empirical nature, Ξ having no direct physical significance and acting only as a correction introduced to simplify the otherwise prohibitively difficult mathematics. There is no obvious method of calculating Ξ from the microscopic properties of the ions and water molecules.

Finally we may mention the assumption of incomplete dissociation made by Nernst† to account for the heats of dilution of electrolytic solutions and the various experiments which seem to justify it and which find the degree of dissociation. In this paper, however, we shall not consider the question of incomplete dissociation of strong electrolytes but shall return to it on a later occasion.

2—PLAN OF THE PAPER

The success of the empirical methods of Hückel and Bonino suggests that the true form of the modified "dielectric constant" derived theoretically should enable a satisfactory extension of the theory of electrolytes to stronger concentrations. We assume a force of a more general type than (1.1), viz.,‡

$$F_{ij} = \varepsilon_i\varepsilon_j/Dr_{ij}^2 - \frac{\partial E_{ij}}{\partial r_{ij}}, \tag{2.1}$$

---

* Bonino, Centola, and Rolla, 'Mem. R. Accad. Ital.,' vol. 4, pp. 415, 445, 465 (1933); 'R.C. Accad. Lincei,' vol. 18, p. 145 (1933); Brüll, 'Gazz. chim. ital.,' vol. 64, pp. 261, 270 (1934).

† 'Z. Elektrochem.,' vol. 33, p. 428 (1927). *See also* Falkenhagen, *op. cit.*

‡ The term $\partial E_{ij}/\partial r_{ij}$ may be regarded as representing the "specific interaction" of the two ions. The treatment given here bears some resemblance to that developed by Guggenheim who combines Brönsted's theory of specific interaction with that of Debye and Hückel; 'Phil. Mag.,' vol. 19, p. 588 (1935).

where the team $\partial E_{ij}/\partial r_{ij}$ acts as a correction, to be evaluated by a direct examination of the close approach of the two ions. Making use of the Gibbs's phase-integral, it is found possible to replace (2.1) by an "effective" force

$$F_{ij} = \epsilon_i \epsilon_j / (D - \delta) r_{ij}^2, \tag{2.2}$$

where δ,* derived from E_{ij}, depends on the temperature and concentration. It is then shown that F has the functional forms

$$F = kTG\,[T(D - \delta)V^{1/3}] = NkT\phi\,[N^{1/3}/T(D - \delta)V^{1/3}], \tag{2.3}$$

where k is Boltzmann's constant, N is the total number of ions in the solution, and V is the volume. We assume that the electrolyte is completely dissociated. The derivations of the relations (2.2) and (2.3) are given in § 3.

In § 4 and § 5 a more detailed analysis is given of Kramers's statistical treatment. It is shown that the various assumptions made by him are very plausible, although rigorous proofs are still partly lacking. In § 6 we introduce the modifications in Kramers's treatment when the force law (2.2) is used in place of (1.1). It is shown that Kramers's method can be extended to stronger concentrations provided the correction term δ is known.

In § 7, the method of finding δ is illustrated by choosing a particular expression for E_{ij}. It is seen that the deviation of Kramers's results from experiment may be attributed to the ionic association at greater concentrations. The actual form of E_{ij} as obtained by examination of the saturation effects and the forces proper between the ions is not treated here but will be considered in a further communication. We conclude with a discussion of the method developed in this paper.

In the Appendix an examination is made of the generalized canonical ensemble used by Kramers. We see that, as a result of the introduction of such an ensemble, there will exist a correction term to his expression for F.

3—Development of the Functional Form

$$F = NkT\phi\,[N^{1/3}/T(D - \delta)V^{1/3}].$$

We shall consider a solution containing ions of the ith kind, each of charge ϵ_i, $i = 1, 2, \ldots, s$, so that $N = \sum_{i=1}^{s} N_i$. Then the phase-integral gives for the free energy

---

* *See* footnote †, p. 535, for the interpretation of δ.

$$e^{-F/kT}\,V^{N} = \int \dots \int e^{-E/kT}\,dx_1\,dy_1\,dz_1 \dots dz_N, \tag{3.1}$$

where, using (2.1),

$$E = \tfrac{1}{2} \sum_{i,j=1}^{N}{}' \,(\varepsilon_i\varepsilon_j/Dr_{ij} + E_{ij}).^{*} \tag{3.2}$$

Here $x_1, y_1, z_1, \dots, z_N$, are the co-ordinates of the ions and E denotes the the instantaneous energy associated with the forces between the ions.

Differentiating (3.1) partially with respect to T, we obtain[†]

$$F - T\frac{\partial F}{\partial T} = \bar{E} - T\,\overline{\frac{\partial E}{\partial T}}, \tag{3.3}$$

where

$$\left.\begin{aligned} \bar{E} &= \left[\int \dots \int E e^{(F-E)/kT}\,dx_1\,dy_1\,dz_1 \dots dz_N\right]\Big/ V^N = \tfrac{1}{2}\sum_{i,j=1}^{N}{}' \,\frac{\overline{\varepsilon_i\varepsilon_j}}{Dr_{ij}} + \bar{\gamma} \\ \bar{\gamma} &= \tfrac{1}{2}\sum_{i,j=1}^{N}{}' \,\bar{E}_{ij}^{\ddagger} \end{aligned}\right\}, \tag{3.4}$$

and

$$\overline{\frac{\partial E}{\partial T}} = \tfrac{1}{2}\sum_{i,j=1}^{N}{}' \,\frac{\overline{\varepsilon_i\varepsilon_j}}{r_{ij}}\left(-\frac{1}{D^2}\frac{dD}{dT}\right) + \overline{\frac{\partial \gamma}{\partial T}}. \tag{3.5}$$

Here $\overline{\frac{\partial E}{\partial T}}$, $\bar{\gamma}$ and $\overline{\frac{\partial \gamma}{\partial T}}$ are defined similarly to $\bar{E}$.

Substituting (3.4) and (3.5) into (3.3), we obtain[§]

$$F - T\frac{\partial F}{\partial T} = \bar{E}\left(1 + \frac{T}{D}\frac{dD}{dT}\right) - \frac{T}{D}\,\overline{\frac{\partial}{\partial T}(D\gamma)}. \tag{3.6}$$

Now

$$\begin{aligned} \frac{\partial F}{\partial \varepsilon_i} &= \left[\int \dots \int \frac{\partial E}{\partial \varepsilon_i} e^{(F-E)/kT}\,dx_1\,dy_1\,dz_1 \dots dz_N\right]\Big/ V^N \\ &= \overline{\frac{\partial E}{\partial \varepsilon_i}} = \bar{\psi}_i + \overline{\frac{\partial \gamma}{\partial \varepsilon_i}}, \end{aligned} \tag{3.7}$$

[*] We use the symbol $\sum'_{i,j,\dots,m}$ when in the summation over $i, j, \dots, m$, $i \neq j \dots \neq m$.

[†] The treatment is similar to that given by Halpern, *loc. cit.*

[‡] We need not assume that γ consists of a sum of terms E_{ij}, implying linear superposition of the fields, but simply treat γ as a whole.

[§] If we put $\bar{\gamma} \equiv 0$ in (3.6) and use the Gibbs-Helmholtz formula for the total internal energy $\Delta U = F - T(\partial F/\partial T)_V$, we then obtain Bjerrum's formula for heat of dilution

$$\Delta U = \bar{E}\left(1 + \frac{T}{D}\frac{dD}{dT}\right).$$

as is obtained from (3.2), writing

$$\bar{\psi}_i = \sum_{j=1}^{N} \overline{\varepsilon_j / D r_{ij}}.$$

Then

$$\bar{E} = \tfrac{1}{2} \sum_{i=1}^{N} \varepsilon_i \bar{\psi}_i + \bar{\gamma} = \tfrac{1}{2} \sum_{i=1}^{N} \varepsilon_i \frac{\partial F}{\partial \varepsilon_i} - \tfrac{1}{2} \sum_{i=1}^{N} \varepsilon_i \overline{\frac{\partial \gamma}{\partial \varepsilon_i}} + \bar{\gamma}. \tag{3.8}$$

Substituting (3.8) into (3.6), we obtain a differential equation for F,

$$\sum_{i=1}^{N} \varepsilon_i \frac{\partial F}{\partial \varepsilon_i} \left(1 + \frac{T}{D} \frac{dD}{dT}\right) + 2T \frac{\partial F}{\partial T} = 2F + \chi(T, V, \varepsilon_i), \tag{3.9}$$

where

$$\chi(T, V, \varepsilon_i) = \frac{2T}{D} \overline{\frac{\partial}{\partial T}(D\gamma)} + \left(1 + \frac{T}{D} \frac{dD}{dT}\right)\left(\sum_{i=1}^{N} \varepsilon_i \overline{\frac{\partial \gamma}{\partial \varepsilon_i}} - 2\bar{\gamma}\right). \tag{3.10}$$

The subsidiary equations of (3.9) form the simultaneous set

$$\frac{d}{dT} \varepsilon_i(T) = \varepsilon_i(T)\left(1 + \frac{T}{D} \frac{dD}{dT}\right) \Big/ 2T, \quad i = 1, \ldots, N, \textit{ i.e.}, \ \varepsilon_i(T) = \sqrt{\kappa_i T D},$$

$$\begin{aligned} \frac{d}{dT} F(T) &= [2F(T) + \chi(T, V, \varepsilon_i(T))]/2T \\ &= [2F(T) + \chi(T, V, \sqrt{\kappa_i T D})]/2T, \end{aligned} \tag{3.11}$$

i.e.,

$$\frac{F(T)}{T} - \tfrac{1}{2} \int^{T} \frac{\chi(T, V, \sqrt{\kappa_i T D})}{T^2} \, dT = \kappa_{N+1},$$

where the κ_i's are constants of integration. Hence the general solution of (3.9) gives

$$F = kTG(\varepsilon_i^2 / kTDV^{1/3}) + H(T, V, \varepsilon_i), \tag{3.12}$$

where

$$H(T, V, \varepsilon_i) = \frac{T}{2} \int^{T} \frac{\chi(T, V, \sqrt{\kappa_i T D})}{T^2} \, dT, \tag{3.13}$$

and we substitute $\kappa_i = \varepsilon_i^2 / TD$ after integration. Here G signifies an arbitrary function of the N arguments $\varepsilon_i^2 / kTDV^{1/3}$, $kV^{1/3}$ being introduced to make the arguments dimensionless.

If we now assume (1.1) to hold, then $\bar{\gamma} \equiv 0$ and we obtain

$$F = kTG(\varepsilon_i^2 / kTDV^{1/3}). \tag{3.14}$$

It follows that if (2.2) is used, giving

$$E = \tfrac{1}{2} \sum_{i,j=1}^{N}{}' \varepsilon_i \varepsilon_j / (D - \delta) r_{ij}, \tag{3.15}$$

the dimensional equation for F reads*

$$F = kTG\,(\varepsilon_i^2/kT\,(D - \delta)\,V^{\frac{1}{3}}), \tag{3.16}$$

since we need only replace kTD by $kT\,(D - \delta)$ in the integrand in (3.1). It is to be noted that (3.14) and (3.16) have the same functional form of their respective arguments.

Equating of (3.12) and (3.16) yields the relations from which δ is to be deduced,

$$G\,(\xi_1) = G\,(\xi) + H\,(T, V, \varepsilon_i)/kT, \tag{3.16}$$

where we consider $G\,(\xi_1)$ as a function of $\xi_1 = T\,(D - \delta)\,V^{\frac{1}{3}}$ on the left and $G\,(\xi)$ one of $\xi = TDV^{\frac{1}{3}}$ on the right. The introduction of δ† in (3.15) implies that we are replacing the quasi-canonical distribution in (3.1) by a different distribution, such that the same value for the free energy is obtained. This new type of distribution will have the characteristic steep maximum at the same value of $E = \bar{E}$ as in (3.1), and hence is legitimate provided the number of ions N is large enough.

It will be shown in § 6 that the Debye-Hückel expression for the free energy F_0 is valid at limiting concentrations ($V \to \infty$), giving

$$F_0 = -\sum_{i=1}^{s} \frac{N_i z_i^2}{3} \frac{\varepsilon^2 \kappa}{D}, \qquad \textit{i.e.,} \qquad G_0\,(\xi) = -\tfrac{2}{3} A N^{\frac{3}{2}} \xi^{-\frac{3}{2}}\;. \tag{3.17}$$

Here $\varepsilon_i = z_i \varepsilon$, ε being the electronic charge and z_i the valency of the ions of the ith kind, $\kappa^2 = \dfrac{4\pi\varepsilon^2}{DkTV} \sum_{i=1}^{s} N_i z_i^2$ and $A = \pi^{\frac{1}{2}}\,\sigma^3/k^{\frac{3}{2}}$, introducing $\sigma^2 = \varepsilon^2 \sum_{i=1}^{s} N_i z_i^2/N$, a constant if the composition of the solution does not vary. Then (3.16) yields as a first approximation to δ

$$-\tfrac{2}{3} A N^{\frac{3}{2}} \xi_1^{-\frac{3}{2}} = -\tfrac{2}{3} A N^{\frac{3}{2}} \xi^{-\frac{3}{2}} + H\,(T, V, \varepsilon_i),$$

or

$$\frac{\delta}{D} = H\,(T, V, \varepsilon_i)/AN^{\frac{3}{2}}\xi^{-\frac{3}{2}} = -\frac{H\,(T, V, \varepsilon_i)}{A}\left(\frac{TDV^{\frac{1}{3}}}{N}\right)^{\frac{3}{2}}, \tag{3.18}$$

---

* *Cf.* Halpern, *loc. cit.*

† No mechanism is introduced for the interpretation of δ. It may be regarded as defined by the relation

$$\int \cdots \int \exp.\left[-\frac{1}{kT}\sum_{i,j=1}^{N}{}' \frac{(1 - e^{-\lambda h_{ij}})}{(D - \delta)\,r_{ij}}\right] dx, dy, dz, \ldots, dz_N = e^{-F/kT}V^N = f_1\left(\frac{N_i}{V}, T\right), \text{ say}$$

where λh_{ij} is introduced in (3.21). Assuming the existence of a solution of this integral equation for $D - \delta$, we have that $D - \delta$ is of the form $D - \delta = f_2\left(\dfrac{N_i}{V}, T\right)$.

putting $(1-\delta/\mathrm{D})^{-\frac{3}{2}} \simeq 1 + \frac{3}{2}\frac{\delta}{\mathrm{D}}$. Better approximations to δ at higher concentrations will be indicated in § 6.

Evidently the evaluation of $\mathrm{H}(\mathrm{T}, \mathrm{V}, \varepsilon_i)$ reduces to that of $\bar{\gamma}$, $\overline{\frac{\partial\gamma}{\partial\mathrm{T}}}$, and $\overline{\frac{\partial\gamma}{\partial\varepsilon_i}}$. Since the contribution $\partial\mathrm{E}_{ij}/\partial r_{ij}$ in (2.1) represents a short-range force, we can find $\bar{\gamma}$ by successive approximations, by treating ionic encounters of increasing orders, as is done in finding the virial for an imperfect gas. This gives $\bar{\gamma}$ in the form $\mathrm{A}_1/\mathrm{V} + \mathrm{A}_2/\mathrm{V}^2 + \ldots$, where A_1, A_2, ..., are functions of the temperature and concentration. Similar forms should be obtained for $\overline{\frac{\partial\gamma}{\partial\varepsilon_i}}$ and $\overline{\frac{\partial\gamma}{\partial\mathrm{T}}}$. It is necessary to treat encounters of order greater than two by some method such as that developed for $\bar{\mathrm{E}}$ by Ursell[*] and applied by the author.[†] In § 7 we shall only evaluate the coefficient A_1, and no investigation of the higher terms will be attempted in this paper.

Since the free energy must be proportional to the total number of ions, if surface phenomena are ignored, it follows that F is of the form $\mathrm{N}\phi(\mathrm{N}/\mathrm{V})$.[‡] Comparing this with (3.16), we derive the dimensional equations

$$\mathrm{F} = k\mathrm{TG}\left[\mathrm{T}(\mathrm{D}-\delta)\mathrm{V}^{\frac{1}{3}}\right] = \mathrm{N}k\mathrm{T}\phi\left[\mathrm{N}^{\frac{1}{3}}/\mathrm{T}(\mathrm{D}-\delta)\mathrm{V}^{\frac{1}{3}}\right]. \tag{3.19}$$

When the expression (3.2) is used for E, the phase-integral in (3.1) may be assumed to be convergent, but it becomes divergent when (3.2) is replaced by (3.15). Adopting the device of Kramers, we choose the energy function

$$\mathrm{E}' = \tfrac{1}{2}\sum_{i,j=1}^{\mathrm{N}}{}' \varepsilon_i\varepsilon_j g_{ij}, \tag{3.20}$$

---

[*] 'Proc. Camb. Phil. Soc.,' vol. 23, p. 685 (1927).

[†] 'Proc. Roy. Soc.,' A, vol. 146, p. 597 (1934). The equation on p. 612 should read

$$\Delta^2\zeta \simeq \frac{8\pi n_i\varepsilon^2}{\mathrm{D}k\mathrm{T}}\left[\cosh\frac{\varepsilon}{k\mathrm{T}}(\psi_1+\psi_2)\right]\zeta + \frac{8\pi n_i\varepsilon}{\mathrm{D}}\left[\sinh\frac{\varepsilon\psi_1}{k\mathrm{T}}\left(\cosh\frac{\varepsilon\psi_2}{k\mathrm{T}}-1\right) + \sinh\frac{\varepsilon\psi_2}{k\mathrm{T}}\left(\cosh\frac{\varepsilon\psi_1}{k\mathrm{T}}-1\right)\right],$$

putting

$$\sinh(\varepsilon\zeta/k\mathrm{T}) \simeq \varepsilon\zeta/k\mathrm{T} \quad \text{and} \quad \cosh(\varepsilon\zeta/k\mathrm{T}) \simeq 1.$$

As a result of the assumptions and approximations made in this paper the final numerical results should be regarded as qualitative only.

[‡] This follows from the form $\sum_{i=1} \mathrm{N}_i\mathrm{K}(\mathrm{N}_i/\mathrm{V})$ given by Halpern, *loc. cit.*, if we assume the composition of the solution is constant.

where

$$g_{ij} = (1 - e^{-\lambda h_{ij}})/(\mathrm{D} - \delta)\, r_{ij}, \tag{3.21}$$

and

$$h_{ij}^2 = \frac{r_{ij}^2}{4\mathrm{R}_{ij}^2} = \frac{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}{(x_i + x_j)^2 + (y_i + y_j)^2 + (z_i + z_j)^2}, \quad \mathrm{R}_{ij} \sim \mathrm{V}^{\frac{1}{3}}.^*$$

We choose λ so large that (3.20) and (3.16) are practically equal for all configurations of the ions. Under these circumstances $1/g_{ij}$ and hence $\varepsilon_i^2/k\mathrm{T}(\mathrm{D} - \delta)$ retain the dimensions of a length, as is essential in our proof.

4—EXAMINATION OF THE SECULAR EQUATION IN KRAMERS'S TREATMENT

In evaluating the Gibbs's phase-integral, Kramers employed a generalized ensemble in which, for a given configuration of the N ions, there correspond N^{N} samples. These are so chosen that for given charges of the ions 1, 2, ..., $k - 1$, $k + 1$, ..., N, there are N_i samples in which the kth ion carries the charge $\varepsilon_i = z_i\varepsilon$. In place of (3.1), we now have

$$\mathrm{V}^{\mathrm{N}} e^{-\mathrm{F}/k\mathrm{T}} = \int \dots \int \mathrm{N}^{-\mathrm{N}}\, \Sigma\, e^{-\mathrm{E}'/k\mathrm{T}}\, dx_1\, dy_1\, dz_1 \dots dz_{\mathrm{N}}, \tag{4.1}$$

where Σ denotes the summation over the N^{N} samples corresponding to a given configuration of the ions. A discussion of this special type of ensemble is given in the Appendix.

Kramers considers the relation (3.20) as representing a quadratic function of the N variables $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{\mathrm{N}}$, and imagines it to be given the canonical form by a transformation to principal axes,

$$\sum_{i,j=1}^{\mathrm{N}}{}' \varepsilon_i \varepsilon_j g_{ij} = \sum_{m=1}^{\mathrm{N}} b_m y_m^2. \tag{4.2}$$

We shall denote by $\Delta_r(\lambda, \mu_{kl})$, the determinant of the rth order, the diagonal terms of which are all equal to λ, the other terms being given by μ_{kl} $(k, l = 1, 2, \dots, r,\ k \neq l)$. Then the quantities b_m, which depend only on the ionic configuration, are the roots of the secular equation

$$\Delta_{\mathrm{N}}(b_1, -g_{ij}) \equiv \sum_{r=0}^{\mathrm{N}} a_r b^{\mathrm{N}-r} = 0, \qquad (a_0 = 1,\ a_1 = 0), \tag{4.3}$$

where

$$a_r = \frac{1}{r\,!} \sum_{i_1, i_2, \dots, i_r = 1}^{\mathrm{N}}{}' \begin{vmatrix} 0, & -g_{i_1 i_2}, & \dots, & -g_{i_1 i_r} \\ \dots & \dots & \dots & \dots \\ -g_{i_r i_1}, & \dots, & -g_{i_r i_{r-1}}, & 0 \end{vmatrix} = \frac{1}{r\,!} \sum_{i_1, i_2, \dots, i_r = 1}^{\mathrm{N}}{}' \Delta_r(0, -g_{\rho\sigma}),$$

$$\rho, \sigma = i_1, i_2, \dots, i_r. \tag{4.4}$$

---

* We shall denote order of magnitude by ~ or O (...).

We first examine the distribution of values of the b_m's. From the recurrence relation $a_0 s_{n-1} + a_1 s_{n-2} + \dots + a_{n-2} s_1 + (n-1)\, a_{n-1} = 0$ the values of

$$s_k = \sum_{m=1}^{N} b_m{}^k,$$

the sum of the kth powers of the roots ($k = 1, . . 2, \dots,$) are easily obtained. Denoting the average value of the N quantities $b_m{}^k$ ($m = 1, 2, .., \mathrm{N}$) for a given ionic configuration by $\overline{b_m{}^k}$, we have

$$s_1 = 0, \quad i.e., \ \overline{b_m} = 0. \tag{4.5}$$

$$s_2 = -2a_2 = -\sum_{i,j=1}^{N}{}' \begin{vmatrix} 0, & -g_{ij} \\ -g_{ji}, & 0 \end{vmatrix} = -\sum_{i,j=1}^{N}{}' \Delta_2(0, \ -g_{ij}).$$

Thus

$$s_2 \sim \mathrm{N}^2 g^2, \ i.e., \ \overline{b_m{}^2} \sim \mathrm{N} g^2, \tag{4.6}$$

where $g = \overline{g_{ij}}$ is the average value of the $\mathrm{N}(\mathrm{N}-1)/2$ quantities g_{ij} for a given ionic configuration. We assume here that $(\overline{g_{ij}})^2 \sim \overline{g_{ij}{}^2}$.

$$s_3 = -3a_3 = -\tfrac{1}{2} \sum_{i,j,k=1}^{N}{}' \Delta_3(0, -g_{\rho\sigma}) \sim \mathrm{N}^3 g^3, \quad \rho, \sigma = i, j, k,$$
$$i.e., \ \overline{b_m{}^3} \sim \mathrm{N}^2 g^3. \tag{4.7}$$

$$s_4 = 2a_2{}^2 - 4a_4 = \tfrac{1}{2}\Big[\sum_{i,j=1}^{N}{}' \Delta_2(0, -g_{ij})\Big]^2 - \tfrac{1}{6} \sum_{i,j,k,l=1}^{N}{}' \Delta_4(0, -g_{\rho\sigma})$$
$$\sim \mathrm{N}^4 g^4, \quad \rho, \sigma = i, j, k, l, \ i.e., \ \overline{b_m{}^4} \sim \mathrm{N}^3 g^4, \tag{4.8}$$

with corresponding expressions for the higher powers. We assume in (4.7) that $\overline{g_{ij} g_{jk} g_{kl}} \sim (\overline{g_{ij}})^3$, and similar relations in (4.8).

To illustrate the evidently unusual distribution of values of the b_m's, we replace (4.3) by the hypothetical equation

$$\Delta_{\mathrm{N}}(b, -g) \equiv \sum_{r=0}^{N} a'_r b^{\mathrm{N}-r} \equiv (b+g)^{\mathrm{N}} \{b - (\mathrm{N}-1)\, g\} = 0, \tag{4.9}$$

where

$$a'_0 = 1, \ a'_1 = 0, \ a'_r = \binom{\mathrm{N}}{r} \Delta_r(0, -g) = \binom{\mathrm{N}}{r} (-g)^r \Delta_r(0, 1). \tag{4.10}$$

To prove the identity in (4.9), we use the recurrence relation

$$\Delta_r(0, 1) + 2\Delta_{r-1}(0, 1) + \Delta_{r-2}(0, 1) = 0,$$

from which, remembering that

$$\Delta_1(0, 1) = 0, \quad \Delta_2(0, 1) = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1,$$

we obtain

$$\Delta_r(0, 1) = (-1)^{r-1}(r-1),$$

so that

$$a'_r = (-1)^{2r-1}(r-1)g^r.$$

Hence the identity in (4.9) is true. Here, then, one root of (4.9) has the value $(N-1)g$ and the other $N-1$ roots are all equal to $-g$. Hence $s_1 = 0$, $s_k \sim N^k g^k$, $(k = 2, 3, \ldots)$, similar to the actual values of s_k.

It has only been possible to indicate that the true values of the b_m's probably have a similar distribution to those of (4.9) and further analysis is desirable. If we treat the a_r's as thermodynamic quantities, so that for the majority of the ionic configurations $|a_r - \overline{a_r}| \ll |\overline{a_r}|$, $\overline{a_r}$ being the average of a_r over all configurations, then we may replace a_r by $\overline{a_r}$ in (4.3). Now $\Delta_r(0, -1) = -(r-1)$ gives the excess of negative elements over the positive elements in the expansion of $\overline{\Delta_r}(0, -g_{\rho\sigma})$. Hence, if all these elements are of the same order of magnitude, then $\overline{a_r}$ is very probably negative for all r (except $a_0 = 1$), leading to one positive root and $N-1$ negative roots in (4.3). This suggests that for the majority of the ionic configurations the b_m's are almost negative. To reconcile a distribution of b_m's, almost all of which are negative, with the values $\overline{b_m} = 0$, $\overline{b_m{}^k} \sim N^{k-1}g^k$, $k = 2, 3, \ldots$, it appears necessary to assume the negative b_m's $(\sim -g)$ are all small, and there is one very large positive b_m $(\sim Ng)$.

The purpose of this investigation is to show that all the negative b_m's satisfy the condition

$$\frac{1}{\sigma^2} + \frac{b_m}{kT} > 0. \tag{4.11}$$

A tentative distribution function $f(b)$ for the b_m's may be obtained by use of Pearson's method,* assuming the moments $\overline{b_m} = 0$, $\overline{b_m{}^2} = Ng^2$, $\overline{b_m{}^3} = N^2g^3$ as given. Then $f(b)$ is the solution of the differential equation

$$\frac{1}{f}\frac{df}{db} = \frac{b-a}{c_0 + c_1 b}, \tag{4.12}$$

where

$$a = c_1 = -N^2g^3/2Ng^2 = -Ng/2, \quad c_0 = -Ng^2,$$

i.e.,

$$f(b) = A[b + 2g]^{-1+4/N} e^{-2b/Ng} \simeq Ae^{-2b/Ng}/(b+2g), \tag{4.13}$$

where b ranges from $-c_0/c_1 = -2g$ to ∞, and A is a constant determined by the condition $\int_{-2g}^{\infty} f(b)\,db = 1$. For a uni-univalent electrolyte $z_1 = -z_2 = 1$, $N_1 = N_2 = \frac{N}{2}$, and the condition (4.11) becomes $b_m > O(-2\cdot 10^5)$, whereas the lower limit of $b = -2g \simeq (\overline{1/Dr_{ij}}) \sim 10^{-2}$, $(V \sim 1)$. Hence we shall assume for the present that the condition (4.11) is satisfied for all the b_m's. It is possible, however, that some of the b_m's may violate (4.11), and a modification in the analysis in § 6 would be necessary.

---

* *See e.g.*, Jordan, " Statistique Mathématique," p. 244.

5—A MORE DETAILED ANALYSIS OF KRAMERS'S DISTRIBUTION FUNCTION FOR y_m

Returning to the transformation (4.2), we have that the y_m's are related to the ε_k's by

$$y_m = \sum_{k=1}^{N} \gamma_{mk}\, \varepsilon_k, \tag{5.1}$$

the coefficients γ_{mk} obeying the orthogonality relations

$$\sum_{k=1}^{N} \gamma^2_{mk} = 1, \qquad \sum_{k=1}^{N} \gamma_{mk}\, \gamma_{m'k} = 0 \qquad (m \neq m'). \tag{5.2}$$

Assuming the quantities $|\gamma_{mk}| \ll 1$ (we shall show $|\gamma_{mk}| \sim 1/\sqrt{N}$ shortly), we follow Kramers and write the summation in the integrand of (4.1) in the form of an N-tuple integral over the values of the y_m's. The bounds of y_m are M and $-$ M, where $M = \varepsilon_m \sum_{k=1}^{N} |\gamma_{mk}| \sim \varepsilon_m \sqrt{N}$, ε_m being the maximum charge present, but we may replace these bounds by $-\infty$ and $+\infty$.

The mean value of y_m is

$$\overline{y}_m = \sum_{k=1}^{N} \gamma_{mk}\overline{\varepsilon}_k = 0, \tag{5.3}$$

since

$$\overline{\varepsilon}_k = \varepsilon \sum_{i=1}^{s} N_i z_i / N = 0. \tag{5.4}$$

Using (5.2) and (5.4), the mean value of y_m^2 is

$$\overline{y_m^2} = \sum_{k=1}^{N} \gamma_{mk}^2 \overline{\varepsilon_k^2} = \varepsilon^2 \sum_{i=1}^{s} N_i z_i^2/N = \overline{\varepsilon_k^2} = \sigma^2. \tag{5.5}$$

Also,

$$\overline{y_m y_{m'}} = \sum_{k=1}^{N} \gamma_{mk}\, \gamma_{m'k}\overline{\varepsilon_k^2} = \sigma^2 \sum_{k=1}^{N} \gamma_{mk}\gamma_{m'k} = 0, \tag{5.6}$$

from which Kramers assumes that the y_m's are statistically independent.

However, the relation (5.6) is not sufficient to establish this property, further conditions being necessary.

Kramers now supposes that the chance that y_m lies between y_m and $y_m + dy_m$ is given by a Gauss error function, *i.e.*,

$$h(y_m)\,dy_m = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-y_m{}^2/2\sigma^2\right] dy_m, \tag{5.7}$$

so that the integrand in (4.1) is given by

$$\mathrm{N}^{-\mathrm{N}} \Sigma\, e^{-\mathrm{E}'/k\mathrm{T}} = \prod_{m=1}^{\mathrm{N}} \left[\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}\left(\frac{1}{\sigma^2} + \frac{b_m}{k\mathrm{T}}\right) y_m{}^2\right\} dy_m\right]. \tag{5.8}$$

We shall now show that the assumptions made by Kramers are very plausible. We examine first the values of the quantities γ_{mk}. There exist the N relations*

$$b_m \gamma_{mp} = \sum_{q=1}^{\mathrm{N}} g_{pq}\gamma_{mq} \qquad (m, p = 1, 2, \ldots, \mathrm{N}), \tag{5.9}$$

which for each b_m give N homogeneous equations to solve for N unknowns $\gamma_{m1}, \gamma_{m2}, \ldots, \gamma_{m\mathrm{N}}$. Solving in the usual manner, we neglect the last equation of (5.9) ($p = \mathrm{N}$), and obtain

$$\gamma_{m1} : \gamma_{m2} : \ldots : \gamma_{m\mathrm{N}} = \mathrm{A}_{m1} : \mathrm{A}_{m2} : \ldots : \mathrm{A}_{m\mathrm{N}}, \tag{5.10}$$

where, if $\mathbf{A}_m$ is the matrix of the system of the remaining N − 1 equations, A_{mk} is the determinant formed from $\mathbf{A}_m$, by omitting the kth column of $\mathbf{A}_m$ and multiplying by the factor $(-1)^{1+k}$. Thus γ_{mk} is proportional to

$$\mathrm{A}_{mk} = (-1)^{1+k} \begin{vmatrix} -b_m, & \ldots, & g_{1,k-1}, & g_{1,k+1}, & \ldots, & g_{1\mathrm{N}} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ g_{\mathrm{N}-1,1}, & \ldots, & g_{\mathrm{N}-1,k-1}, & g_{\mathrm{N}-1,k+1}, \ldots, & -b_m, & g_{\mathrm{N}-1,\mathrm{N}} \end{vmatrix}. \tag{5.11}$$

Now if we treat A_{mk} as a thermodynamic quantity,

$$(|\mathrm{A}_{mk} - \bar{\mathrm{A}}_{mk}| \ll |\bar{\mathrm{A}}_{mk}|),$$

and neglect the fluctuations we may replace A_{mk} by $\bar{\mathrm{A}}_{mk}$, its average value taken over all configurations of the ions, and therefore γ_{mk} by $\bar{\gamma}_{mk}$. Now if the ions k and l are of the same type $|\bar{\gamma}_{mk}| = |\bar{\gamma}_{ml}|$, and if they are of different type, we may assume that $|\bar{\gamma}_{mk}| \sim |\bar{\gamma}_{ml}|$.† It follows from (5.2) that $|\gamma_{mk}| \sim 1/\mathrm{N}^{\frac{1}{2}}$ which will hold for all k, $m = 1, 2, \ldots, \mathrm{N}$.

---

* *See* Courant and Hilbert, "Methoden der Mathematischen Physik," 2nd Auflage, p. 20 (1931).

† For a symmetrical type, $|\bar{\gamma}_{mk}| = |\bar{\gamma}_{ml}|$ for all k, l.

We assume the statistical independence of the y_m's for the moment and pass on to consideration of (5.7). In order that (5.7) be the true distribution function it is necessary that

$$\overline{y_m^{2n}} = \frac{(2n)\,!}{2^n n\,!}\,\sigma^{2n}, \qquad \overline{y_m^{2n+1}} = 0. \tag{5.12}$$

We shall proceed to find expressions for $\overline{y_m^p}$ $(p = 1, 2, \ldots)$.

Using (5.1), (5.2), and (5.4), and applying the multinomial theorem,

$$\overline{y_m^4} = \overline{(\sum_{k=1}^{N} \gamma_{mk}\varepsilon_k)^4} = \overline{\varepsilon_k^4} \sum_{k=1}^{N} \gamma_{mk}^4 + \binom{4}{2}\frac{1}{2\,!}\,\overline{\varepsilon_k^2}\,.\,\overline{\varepsilon_l^2} \sum_{k,l=1}^{N}{}' \gamma_{mk}^2\,\gamma_{ml}^2.$$

Now

$$\sum_{k=1}^{N} \gamma_{mk}^4 \sim 1/\mathrm{N}, \qquad \sum_{k,l=1}^{N}{}' \gamma_{mk}^2\,\gamma_{ml}^2 = \sum_{k,l=1}^{N} \gamma_{mk}^2\,\gamma_{ml}^2 - \sum_{k=1}^{N} \gamma_{mk}^4 = 1 - \mathrm{O}\,(1/\mathrm{N}).$$

Hence

$$\frac{\left|\overline{y_m^4} - \dfrac{4!}{2^2}\dfrac{1}{2!}\,\overline{\varepsilon_k^2}\,.\,\overline{\varepsilon_l^2}\right|}{\dfrac{4!}{2^2}\dfrac{1}{2!}\,\overline{\varepsilon_k^2}\,.\,\overline{\varepsilon_l^2}} = \frac{\left|\overline{y_m^4} - \dfrac{4!}{2^2}\dfrac{1}{2!}\,\sigma^4\right|}{\dfrac{4!}{2^2}\dfrac{1}{2!}\,\sigma^4} \sim \frac{1}{\mathrm{N}}\,.$$

Similarly

$$\overline{y_m^6} = \overline{\varepsilon_k^6} \sum_{k=1}^{N} \gamma_{mk}^6 + \binom{6}{2}\,\overline{\varepsilon_k^4}\,.\,\overline{\varepsilon_l^2} \sum_{k,l=1}^{N}{}' \gamma_{mk}^4\gamma_{ml}^2 + \binom{6}{3}\frac{1}{2\,!}\,\overline{\varepsilon_k^3}\,.\,\overline{\varepsilon_l^3} \sum_{k,l=1}^{N}{}' \gamma_{mk}^3\gamma_{ml}^3$$
$$+ \frac{6\,!}{(2\,!)^3}\,\frac{1}{3\,!}\,\overline{\varepsilon_j^2}\,.\,\overline{\varepsilon_k^2}\,.\,\overline{\varepsilon_l^2} \sum_{j,k,l=1}^{N}{}' \gamma_{mj}^2\gamma_{mk}^2\gamma_{ml}^2.$$

Here

$$\sum_{k=1}^{N} \gamma_{mk}^6 \sim 1/\mathrm{N}^2, \qquad \sum_{k,l=1}^{N}{}' \gamma_{mk}^4\gamma_{ml}^2 \sim 1/\mathrm{N}, \qquad \left|\sum_{k,l=1}^{N}{}' \gamma_{mk}^3\gamma_{ml}^3\right| < 1/\mathrm{N},$$

$$\sum_{j,k,l=1}^{N}{}' \gamma_{mj}^2\gamma_{mk}^2\gamma_{ml}^2 = \sum_{j,k,l=1}^{N} \gamma_{mj}^2\gamma_{mk}^2\gamma_{ml}^2 - \frac{3\,!}{2\,!} \sum_{k,l=1}^{N}{}' \gamma_{mk}^4\gamma_{ml}^2$$
$$- \sum_{k=1}^{N} \gamma_{mk}^6 = 1 - \mathrm{O}\,(1/\mathrm{N}) - \mathrm{O}\,(1/\mathrm{N}^2).$$

Hence

$$\frac{\left|\overline{y_m^6} - \dfrac{6\,!}{2^3\,3\,!}\,\sigma^6\right|}{\dfrac{6\,!}{2^3\,3\,!}\,\sigma^6} \sim \frac{1}{\mathrm{N}}.$$

We can similarly prove that in general

$$\frac{\left|\overline{y_m^{2n}} - \dfrac{(2n)\,!}{2^n\,n\,!}\,\sigma^{2n}\right|}{\dfrac{(2n)\,!}{2^n\,n\,!}\,\sigma^{2n}} \sim \frac{1}{\mathrm{N}}. \tag{5.13}$$

On the other hand, for an electrolyte of the non-symmetrical type,

$$|\overline{y_m^3}| = |\overline{\varepsilon_k^3} \sum_{k=1}^{N} \gamma_{mk}^3| \leqslant |\overline{\varepsilon_k^3}|/N^{\frac{1}{2}} \sim \sigma^3/N^{\frac{1}{2}},$$

$$|\overline{y_m^5}| = |\overline{\varepsilon_k^5} \sum_{k=1}^{N} \gamma_{mk}^5 + \binom{5}{3} \overline{\varepsilon_k^3} \,.\, \overline{\varepsilon_l^2} \sum_{k,l=1}^{N}{}' \gamma_{mk}^3 \gamma_{ml}^2| < |\overline{\varepsilon_k^5}|/N^{\frac{1}{2}}$$
$$+ \binom{5}{3} |\overline{\varepsilon_k^3}| \, \overline{\varepsilon_l^2}/N^{\frac{1}{2}} \sim \sigma^5/N^{\frac{1}{2}}.$$

For a binary electrolyte of the symmetrical type, $\overline{\varepsilon_k^{2n+1}} = 0$ and hence $\overline{y_k^{2n+1}} = 0$. Thus in general,

$$|\overline{y_m^{2n+1}}| \leqslant O\,(\sigma^{2n+1}/N^{\frac{1}{2}}) \ll \sigma \,.\, \overline{y_m^{2n}}, \tag{5.14}$$

so that we may assume the second of the relations (5.13). It follows that the Gauss error distribution function may be used and thus the relation (5.8) is valid.*

---

* It is not necessary to know $\overline{y_m^{2n+1}}$ in order to show that (5.8) is correct. If the moments $\overline{y_m^n}$ of $h\,(y_m)$ are given and y_m is assumed to range from $-\infty$ to $+\infty$, then $h\,(y_m)$ is expressed in series in Hermitian polynomials as

$$h\,(y_m) = \frac{1}{\sigma\sqrt{2\pi}}\left[c_0 + \frac{c_1}{1!}\,H_1\,(y_m/\sigma) + \frac{c_2}{2!}\,H_2\,(y_m/\sigma) + \ldots\right] e^{-y_m^2/2\sigma^2}, \tag{I}$$

with y_m/σ as argument. Here $H_n\,(x)$ is defined by

$$e^{xt-\frac{1}{2}t^2} = \sum_{n=0}^{\infty} H_n\,(x)\, t^n/n\,!,$$

with the orthogonality conditions

$$\int_{-\infty}^{+\infty} H_n\,(x)\, H_{n'}\,(x)\, dx = 0 \text{ for } n \neq n',$$
$$= n\sqrt{2\pi} \text{ for } n = n'. \tag{II}$$

Using these relations, we see that

$$c_n = \int_{-\infty}^{+\infty} H_n\,(y_m/\sigma)\, h\,(y_m)\, dy_m. \tag{III}$$

In particular $c_0 = 1$, $c_1 = c_2 = 0$. Substituting the Hermitian polynomial for $n = 2p$,

$$H_{2p}\,(x) = x^{2p} - \frac{2p\,(2p-1)}{1\,!}\frac{x^{2p-2}}{2} + \frac{2p\,(2p-1)\,(2p-2)\,(2p-3)}{2\,!}\frac{x^{2p-4}}{2^2}$$
$$+ \ldots + (-1)^p \frac{(2p)\,!}{p\,!}\frac{1}{2^p},$$

in (III) and employing (II) and the first relation in (5.12), it is easily seen that $c_{2p} = 0$. Thus (5.8) becomes

$$h\,(y_m) = \left[1 + \frac{c_3}{3\,!}\,H_3\,(y_m/\sigma) + \frac{c_5}{5\,!}\,H_5\,(y_m/\sigma) + \ldots\right] \frac{1}{\sigma\sqrt{2\pi}}\, e^{-y_m^2/2\sigma^2}. \tag{IV}$$

Remembering that $H_{2n+1}\,(y_m/\sigma)$ is an odd function, the relation (5.8) follows from (IV).

There is left to show the statistical independence of the y_m's. The necessary and sufficient condition for y_m and $y_{m'}$ to be independent is that $(\overline{y_m} = \overline{y_{m'}} = 0)$*

$$r_{p|q} - r_{p|0} \cdot r_{0|q} = 0,$$

where

$$r_{\lambda|\mu} = \overline{y_m{}^{\lambda} y_{m'}{}^{\mu}} / (\overline{y_m{}^2})^{\lambda/2} (\overline{y_{m'}{}^2})^{\mu/2},$$

i.e,

$$[\overline{y_m{}^p y_{m'}{}^q} - \overline{y_m{}^p} \cdot \overline{y_{m'}{}^q}] / [(\overline{y_m{}^2})^{p/2} (\overline{y_{m'}{}^2})^{q/2}] = 0, \tag{5.15}$$

for all positive integral values of p, q. Making use of (5.12), it is readily seen that the denominator of (5.15) is of the order of magnitude of σ^{p+q}. When not both of p, q are given, we have from (5.14) that

$$|\overline{y_m{}^p} \cdot \overline{y_{m'}{}^q}| \leqslant O\,(\sigma^{p+q}/N^{\frac{1}{2}}).$$

The same is true of the other term in the numerator of (5.15) $\overline{y_m{}^p y_{m'}{}^q}$. For example, using (5.4),

$$\overline{y_m y_{m'}{}^2} = \sum_{i,j=1}^{N} \gamma_{mi} \gamma_{m'j}{}^2 \overline{\varepsilon_i \varepsilon_j{}^2} + \sum_{i=1}^{N} \sum_{k,l=1}^{N}{}' \gamma_{mi} \gamma_{m'k} \gamma_{m'l} \overline{\varepsilon_i \varepsilon_k \varepsilon_l} = \sum_{i=1}^{N} \gamma_{mi} \gamma_{m'i}{}^2 \overline{\varepsilon_i{}^3}$$
$$+ \sum_{i,j=1}^{N}{}' \gamma_{mi} \gamma_{m'j}{}^2 \overline{\varepsilon_i \varepsilon_j{}^2} + 2 \sum_{k,l=1}^{N}{}' \gamma_{mk} \gamma_{m'k} \gamma_{m'l} \overline{\varepsilon_k{}^2 \varepsilon_l} + \sum_{i,k,l=1}^{N}{}' \gamma_{mi} \gamma_{m'k} \gamma_{m'l} \overline{\varepsilon_i \varepsilon_k \varepsilon_l},$$

or

$$|\overline{y_m y_{m'}{}^2}| = |\overline{\varepsilon_i{}^3} \sum_{i=1}^{N} \gamma_{mi} \gamma_{m'i}{}^2| \sim \left| \frac{\overline{\varepsilon_i{}^3}}{N} \sum_{i=1}^{N} \gamma_{mi} \right| < |\overline{\varepsilon_i{}^3}|/N^{\frac{1}{2}} \leqslant O\,(\sigma^3/N^{\frac{1}{2}}).$$

When p, q are both even we have

$$|\overline{y_m{}^p y_{m'}{}^q} - \overline{y_m{}^p} \cdot \overline{y_{m'}{}^q}| \leqslant O\,(\sigma^{p+q}/N).$$

For example, by (5.4), the non-vanishing terms in $\overline{y_m y_{m'}{}^2}$ are

$$\overline{y_m{}^2 y_{m'}{}^2} = \sum_{i,j=1}^{N} \gamma_{mi}{}^2 \gamma_{m'j}{}^2 \overline{\varepsilon_i{}^2 \varepsilon_j{}^2} + 2 \sum_{k,l=1}^{N}{}' \gamma_{mk} \gamma_{m'k} \gamma_{ml} \gamma_{m'l} \overline{\varepsilon_k{}^2 \varepsilon_l{}^2}$$
$$= \sum_{i,j=1}^{N}{}' \gamma_{mi}{}^2 \gamma_{m'j}{}^2 \overline{\varepsilon_i{}^2 \varepsilon_j{}^2} + \sum_{i=1}^{N} \gamma_{mi}{}^2 \gamma_{m'j}{}^2 \overline{\varepsilon_i{}^4} + 2 \overline{\varepsilon_k{}^2 \varepsilon_l{}^2} (\sum_{k=1}^{N} \gamma_{mk} \gamma_{m'k})^2$$
$$- 2 \sum_{k=1}^{N} \gamma_{m'k}{}^2 \gamma_{mk}{}^2 \overline{\varepsilon_k{}^2} \cdot \overline{\varepsilon_l{}^2},$$

i.e.,

$$|\overline{y_m{}^2 y_{m'}{}^2} - \overline{\varepsilon_i{}^2} \cdot \overline{\varepsilon_j{}^2}| = |\overline{y_m{}^2 y_{m'}{}^2} - \overline{y_m{}^2} \cdot \overline{y_{m'}{}^2}| = O\,(\sigma^4/N).$$

---

* *See* Borel, "Traité du Calcul des Probabilités et de ses Applications," vol. 1, part 4 (part 2), chap. 3.

We conclude that $|r_{p|q}| \leqslant O(1/N^{\frac{1}{2}})$ of all p, q, and hence the y_m's can be taken to be statistically independent.

6—EVALUATION OF THE FREE ENERGY OF THE ELECTROLYTIC SOLUTION

Using (4.13), the value of (5.8), as derived by Kramers, is

$$N^{-N} \Sigma e^{-E'/kT} = \prod_{m=1}^{N} \left(1 + \frac{\sigma^2 b_m}{kT}\right)^{-\frac{1}{2}} = \left[\left(-\frac{\sigma^2}{kT}\right)^N \Delta_N\left(-\frac{kT}{\sigma^2}, -g_{kl}\right)\right]^{-\frac{1}{2}} \quad (6.1)$$

$$= \Delta_N^{-\frac{1}{2}}(1, \alpha g'_{kl}) = A_N^{-\frac{1}{2}}, \quad \text{say,}$$

where

$$\alpha = \sigma^2/(D - \delta)\, kT = \beta/(1 - \theta(N)) \text{ and } g_{kl} = \frac{g'_{kl}}{D - \delta}. \quad (6.2)$$

Here $\beta = \sigma^2/DkT$ and for convenience we indicate the argument N only in $\theta(N) = \delta/D$, δ being also a function of T and V. We regard D and therefore β as a function of T only. Substituting (6.1) into (4.1), we obtain

$$F = -kT \log \overline{A_N^{-\frac{1}{2}}} = \tfrac{1}{2}kT \log \overline{A_N} = \tfrac{1}{2}kT \log \overline{\Delta_N}(1, \alpha g'_{kl}). \quad (6.3)$$

Here the bar denotes the mean value taken over all configurations of the ions of the volume V and we neglect the fluctuations of A_N, replacing $\overline{A_N^{-\frac{1}{2}}}$ by $\overline{A_N}^{-\frac{1}{2}}$.

We proceed to derive a general expression for $\overline{A_N}$. Differentiating $\overline{A_N}$ partially with respect to β,

$$\frac{\partial \overline{A_N}}{\partial \beta} = \frac{\partial}{\partial \beta}\left[\alpha^N \Delta_N\left(\frac{1}{\alpha}, g'_{kl}\right)\right] = \frac{N}{\alpha}\frac{\partial \alpha}{\partial \beta}\overline{A}_N + \alpha^N \frac{\partial}{\partial \beta}\overline{\Delta}_N\left(\frac{1}{\alpha}, g'_{kl}\right). \quad (6.4)$$

Remembering that the mean values of the subdeterminants of the diagonal terms of $\overline{\Delta}_N\left(\frac{1}{\alpha}, g'_{kl}\right)$ are all equal, we have

$$\frac{\partial}{\partial \beta}\overline{\Delta}_N\left(\frac{1}{\alpha}, g'_{kl}\right) = \frac{\partial}{\partial \beta}\overline{\Delta}_N\left(\frac{1-\theta(N)}{\beta}, g'_{kl}\right) = -\frac{N}{\alpha^2}\frac{\partial \alpha}{\partial \beta}\overline{\Delta}_{N-1}\left(\frac{1-\theta(N)}{\beta}, g'_{kl}\right). \quad (6.5)$$

Retaining the first two terms in the expansion of $\theta(N - 1)$,

$$\begin{aligned}\overline{\Delta}_{N-1}\left(\frac{1-\theta(N-1)}{\beta}, g'_{kl}\right) &= \overline{\Delta}_{N-1}\left(\frac{1-\theta(N)}{\beta} + \frac{1}{\beta}\frac{\partial \theta}{\partial N}, g'_{kl}\right) \\ &= \overline{\Delta}_{N-1}\left(\frac{1-\theta(N)}{\beta}, g'_{kl}\right) + \frac{(N-1)}{\beta}\frac{\partial \theta}{\partial N}\overline{\Delta}_{N-2}\left(\frac{1-\theta(N-1)}{\beta}, g'_{kl}\right) \\ &\quad + \left[\frac{1-\theta(N)}{\beta} + \frac{1}{\beta}\frac{\partial \theta}{\partial N}\right]^{N-1} - \left[\frac{1-\theta(N)}{\beta}\right]^{N-1} \\ &\quad - \frac{(N-1)}{\beta}\frac{\partial \theta}{\partial N}\left[\frac{1-\theta(N)}{\beta} + \frac{1}{\beta}\frac{\partial \theta}{\partial N}\right]^{N-2}. \quad (6.6)\end{aligned}$$

Since $N\dfrac{\partial\theta}{\partial N} \sim 1$ ($\theta \sim 1$ and for dilute solutions $\theta < 1$), the last three terms on the right of (6.6) are of the order of magnitude of

$$[1 - \theta(N)]^{N-1}/\beta^{N-1}.$$

Using the hypothetical case (4.9),

$$\bar{\Delta}_{N-1}\left(\frac{1-\theta(N)}{\beta}, g'_{kl}\right) \sim \sum_{r=0}^{N-1}{}' (-1)^{2r-1}\binom{N-1}{r}(r-1)\,g^r\left[\frac{1-\theta(N)}{\beta}\right]^{N-1-r}, \tag{6.7}$$

with an analogous expression for $\bar{\Delta}_{N-2}$. Since $\beta \sim 10^{-7}$, it follows that we need only retain the first two terms on the right of (6.6), the others being negligible in comparison. This gives us a recurrence relation yielding the series

$$\begin{aligned}\bar{\Delta}_{N-1}\left(\frac{1-\theta(N)}{\beta}, g'_{kl}\right) &= \bar{\Delta}_{N-1}\left(\frac{1-\theta(N-1)}{\beta}, g'_{kl}\right)\\ &\quad - \frac{(N-1)}{\beta}\frac{\partial\theta}{\partial N}\,\bar{\Delta}_{N-2}\left(\frac{1-\theta(N-2)}{\beta}, g'_{kl}\right) + \dots\end{aligned} \tag{6.8}$$

Retaining the first term on the right of (6.8) and substituting the thereby modified form (6.5) into (6.4) we derive

$$\frac{\partial\bar{A}_N}{\partial\beta} = \frac{N}{\alpha}\frac{\partial\alpha}{\partial\beta}(\bar{A}_N - \bar{A}_{N-1}) = \frac{N}{\alpha}\frac{\partial\alpha}{\partial\beta}\frac{\partial\bar{A}_N}{\partial N}, \tag{6.9}$$

to a first approximation. The solution of (6.9) is of the form

$$\bar{A}_N = \phi(\alpha N, V) = \left[\omega\left\{\frac{N^{1/3}\beta}{(1-\theta)V^{1/3}}\right\}\right]^N = \left[\omega\left(\frac{\alpha N^{1/3}}{V^{1/3}}\right)\right]^N, \tag{6.10}$$

where ω^N is a second functional form for A_N obtained from (3.1) and (6.3). On differentiating (6.10) partially with respect to β and N and eliminating ϕ from the two resulting equations we arrive at

$$\tfrac{2}{3}\zeta\frac{d\log\omega}{d\zeta} = \log\omega, \qquad \zeta = \alpha N^{1/3}/V^{1/3}, \tag{6.11}$$

giving

$$\bar{A}_N = \omega^N = \exp[-2K(\beta N)^{3/2}V^{-1/2}(1-\theta)^{-3/2}]. \tag{6.12}$$

Here $K = \tfrac{2}{3}\sqrt{\pi}$, as shown by Kramers. Hence

$$\begin{aligned}F &= \tfrac{1}{2}NkT\log\omega = -\tfrac{2}{3}\sqrt{\pi}\,kT(\beta N)^{3/2}V^{-1/2}(1-\theta)^{-3/2}\\ &= -\sum_{i=1}^{s}\frac{N_i z_i^2}{3}\frac{\varepsilon^2\kappa}{D}(1-\theta)^{-1/2},\end{aligned} \tag{6.13}$$

which corresponds to the Debye-Hückel second approximation.

As noted by Kramers, it follows from (6.12) that for larger concentrations, the ratio between $\bar{A}_N$ and $\bar{A}_{N-1}$ will deviate more and more from unity, so that the approximation in (6.9) is no longer permissible. Proceeding in the manner adopted by Kramers, we have

$$\begin{aligned}\bar{A}_N - \bar{A}_{N-1} &= [\omega(x_N)]^N - [\omega(x_{N-1})]^{N-1} \\ &= [\omega(x_N)]^{N-1}\left[\omega(x_N) - \left(1 - \frac{\omega'(x_N)}{\omega(x)_N}\frac{\partial x_N}{\partial N}\right)^{N-1}\right] \\ &= [\omega(x_N)]^{N-1}\left[\omega(x_N) - \exp\left(-\frac{N\omega'(x_N)}{\omega(x_N)}\frac{\partial x_N}{\partial N}\right)\right], \qquad (6.14)\end{aligned}$$

where $x_N = \alpha^3 N/V$ is considered to be a function of N and $\omega'(x_N) = d\omega(x_N)/dx_N$. Then (6.9) becomes, omitting the argument of ω and ω',

$$\frac{\partial \bar{A}_N}{\partial \beta} = 3N\omega^{N-1} x_N \frac{\partial \alpha}{\partial \beta}\frac{\omega'}{\alpha} = \frac{N\omega^{N-1}}{\alpha}\frac{\partial \alpha}{\partial \beta}\left[\omega - \exp\left(-\frac{N\omega'}{\omega}\frac{\partial x_N}{\partial N}\right)\right], \qquad (6.15)$$

i.e.,

$$3x\omega' = \omega - \exp\left(-\frac{x\omega'}{\omega}\frac{N}{x}\frac{\partial x}{\partial N}\right) = \omega - \exp\left[-\frac{x\omega'}{\omega}\left(1 + \frac{3N}{1-\theta}\frac{\partial \theta}{\partial N}\right)\right], \qquad (6.16)$$

dropping the suffix on x. Here $\frac{3N}{1-\theta}\frac{\partial \theta}{\partial N}$ and hence ω is no longer a function of x alone, the discrepancy arising when $\omega(x_{N-1})$ is expanded in (6.14). To overcome this difficulty we assume T is kept constant so that x is a measure of the concentration, and we write

$$\psi(x) = \frac{3N}{1-\theta}\frac{\partial \theta}{\partial N}.$$

On putting $\theta = 0$, (6.16) reduces to

$$3x_0\omega'_0 = \omega_0 - \exp\left(-\frac{x_0\omega'_0}{\omega_0}\right), \qquad (6.17)$$

where $\omega = \omega_0$, $x = x_0$, $\omega'_0 = \frac{d\omega_0}{dx_0}$ when $\theta = 0$. Substituting $\frac{x_0\omega'_0}{\omega_0} = -t_0$, (6.17) becomes

$$\omega_0 = e^{t_0}/(1 + 3t_0), \qquad (6.18)$$

which on differentiating with respect to x_0 leads to $\left(t'_0 = \frac{dt_0}{dx_0}\right)$

$$-t_0 = \frac{x_0\omega'_0}{\omega_0} = -\frac{(2-3t_0)\,t'_0 x_0}{1+3t_0},$$

with solution

$$K^2 x_0 = \frac{4\pi}{9} x_0 = t_0^2/(1+3t_0)^3 = \frac{4\pi}{9} \beta^3 \frac{N}{V}. \tag{6.19}$$

Here (6.18) and (6.19) are the parametric solution obtained by Kramers.

We now treat (6.16) in a similar fashion. Putting $x\omega'/\omega = -t$, (6.16) becomes

$$\omega = e^{t(1+\psi)}/(1+3t) = e^t e^\chi/(1+3t), \tag{6.20}$$

introducing $\chi = t\psi$. Differentiating logarithmically gives

$$\left(\psi' = \frac{d\psi}{dx}, \qquad t' = \frac{dt}{dx}\right)$$

$$-t = \frac{x\omega'}{\omega} = \psi' t x + (1+\psi) t' x - 3t' x/(1+3t),$$

or

$$\frac{dx}{x} = \frac{(2-3t) - \psi(1+3t)}{t(1+3t)(1+\psi' x)} dt.$$

The solution is

$$K_1^2 x = K_1^2 \frac{\beta^3}{(1-\theta)^3} \frac{N}{V} = \frac{t^2 e^{v(t)}}{(1+3t)^3}, \tag{6.21}$$

where

$$v(t) = \int \left\{ \frac{2-3t}{t(1+3t)} \left[\frac{1}{1+\psi' x} - 1 \right] - \frac{\psi}{t(1+\psi' x)} \right\} dt. \tag{6.22}$$

We assume in (6.20) and (6.22) that $x(t)$ and hence $\psi(t)$, $\chi(t)$ and $v(t)$ are known functions of t. Then the solution given by (6.20) and (6.21) replaces (6.18) and (6.19).

We now derive

$$\log f_i = \frac{1}{kT} \frac{\partial F}{\partial N_i} = \frac{N}{2} \frac{1}{\omega} \frac{d\omega}{dx} \frac{\partial x}{\partial N_i} + \tfrac{1}{2} \log \omega$$

$$= \tfrac{1}{2} [t + \chi - \log(1+3t) + t\,\Phi(t)\,\Psi], \tag{6.23}$$

and

$$1 - g = \frac{V}{NkT} \frac{\partial F}{\partial V} = \frac{t}{2} \left(1 - \frac{3V}{D(1-\theta)} \frac{\partial \delta}{\partial V}\right) \Phi(t). \tag{6.24}$$

To obtain (6.23) and (6.24), we use the relations

$$\left.\begin{aligned}
\frac{x\omega'}{\omega} &= -t\,\Phi(t) = -t \left[1 - \frac{1+3t}{2-3t} \frac{d\chi}{dt}\right] \Big/ \left[1 + \frac{t(1+3t)}{2-3t} \frac{dv}{dt}\right], \\
\frac{\partial x}{\partial V} &= -\frac{x}{V} \left(1 - \frac{3V}{D(1-\theta)} \frac{\partial \delta}{\partial V}\right), \\
\frac{\partial x}{\partial N_i} &= \frac{x}{N} \left[\frac{3N}{D(1-\theta)} \frac{\partial \delta}{\partial N_i} + \frac{3N z_i^2}{\sum_{i=1}^{s} N_i z_i^2} - 2 \right] = \frac{x}{N} \Psi,
\end{aligned}\right\} \tag{6.25}$$

as may be deduced from (6.20) and (6.21), noting that $\theta = \delta/D$. On putting $\theta = 0$, (6.25) and (6.26) reduce to Kramers's expressions

$$\log f_i = \frac{3}{2} t \left(1 - \frac{N z_i^2}{\sum\limits_{i=1}^{s} N_i z_i^2} \right) - \tfrac{1}{2} \log (1 + 3t), \qquad 1 - g = \frac{t}{2}. \quad (6.26)$$

To indicate the modifications produced in the osmotic coefficients, we develop (6.20) and (6.21) further for dilute solutions. According to (3.18), a first approximation to δ is

$$\delta \simeq \frac{\lambda_1 (T)}{A} \left(\frac{T^3 D^5}{V} \right)^{\frac{1}{2}} N^{\frac{1}{2}}, \tag{6.27}$$

so that

$$\frac{3N}{1 - \theta} \frac{\partial \theta}{\partial N} = \tfrac{3}{2} \frac{\delta}{D (1 - \theta)} = \mu_1 (T)\, x^{\frac{1}{2}} + \dots\dots, \tag{6.28}$$

where $\mu_1 (T) = \frac{3}{2} \lambda_1 (T) (TD/\beta)^{\frac{1}{2}}$. Since we shall not evaluate $\lambda_1 (T), \dots$ in this paper, an estimate of δ is obtained from the work of Debye and Hückel. Writing their expression for the electrostatic energy as

$$\bar{E} = - \sum_{i=1}^{s} \frac{N_i z_i^2}{2D} \frac{\varepsilon^2 \kappa}{1 + \kappa a} = - \frac{A N^{\frac{3}{2}} kT}{[D (1 + \kappa a)^{\frac{2}{3}} T V^{\frac{1}{3}}]^{\frac{3}{2}}} = - \frac{A N\, kT}{[(D - \delta) T V^{\frac{1}{3}}]^{\frac{3}{2}}},$$

where a is the " average " radius of the ions, we derive

$$1 - \theta = 1 - \delta/D = (1 + \kappa a)^{\frac{2}{3}}, \tag{6.29}$$

giving to a first approximation

$$\frac{\delta}{D} = - \tfrac{2}{3} \kappa a = - \tfrac{4}{3} \left(\pi \beta \frac{N}{V} \right)^{\frac{1}{2}} a = - \tfrac{4}{3} \sqrt{\pi} \frac{(1 - \theta)^3}{\beta} x^{\frac{1}{2}} a. \tag{6.30}$$

Then

$$\psi (x) \simeq - 2 \sqrt{\pi} (1 - \theta)^{\frac{1}{2}} x^{\frac{1}{2}} a \simeq - \frac{2 \sqrt{\pi} a}{\beta} x^{\frac{1}{2}} = \mu_1 x^{\frac{1}{2}}, \quad \chi = t \mu_1 x^{\frac{1}{2}}. \tag{6.31}$$

Writing $1/(1 + \psi' x) = 1 - \frac{1}{2} \mu_1 x^{\frac{1}{2}}$ and retaining the term $x^{\frac{1}{2}}$ only, we derive for (6.22)

$$v (t) = - \frac{\mu_1}{2} \int \frac{x^{\frac{1}{2}} (4 + 3t)}{t (1 + 3t)} dt. \tag{6.32}$$

Substituting $x^{\frac{1}{2}} = t/K (1 + 3t)^{\frac{1}{2}}$ as a first approximation, (6.32) becomes

$$v_1 (t) = \frac{\mu_1}{3K} \frac{(2 + 3t)}{(1 + 3t)^{\frac{1}{2}}} + A_1, \tag{6.33}$$

where A_1 is a constant. Then (6.22) reads

$$K_1^2 x_1 = \frac{t^2}{(1+3t)^3} \exp\left[\frac{\mu_1}{3K} \frac{(2+3t)}{(1+3t)^{\frac{1}{2}}} + A_1\right]. \qquad (6.34)$$

If we assume that (6.34) must reduce to (6.19) for small t (neglecting all orders higher than that of t^2), then $A_1 = -2\mu_1/3K$, $K_1 = K$.

Continuing this process (*i.e.*, we now use (6.34) for $x^{\frac{1}{2}}$), we obtain for the second approximation

$$v_2(t) = 2(\zeta - 1), \quad \zeta = \exp\left[\frac{\mu_1}{6K} \frac{(2+3t)}{(1+3t)^{\frac{1}{2}}} - \frac{\mu_1}{3K}\right], \qquad (6.35)$$

$$K^2 x_2 = \frac{t^2}{(1+3t)^3} \exp\left[2(\zeta - 1)\right], \qquad (6.36)$$

and for the third

$$v_3(t) = \frac{2}{e}\left\{\left(\log\zeta + \frac{\zeta}{1} + \frac{\zeta^2}{1.2^2} + \frac{\zeta^3}{1.2.3^2} + \ldots\right) - \left(1 + \frac{1}{1.2^2} + \frac{1}{1.2.3^2} + \ldots\right)\right\}, \qquad (6.37)$$

with the corresponding expression for x_3. It was found that (6.35) and (6.37) gave practically the same results, so that there was no need for higher approximations.

Using (6.31) and (6.36), we can now find $x(t)$ and from (6.35) $v(t)$. Further, according to (6.29),

$$\frac{3V}{D(1-\theta)} \frac{\partial\delta}{\partial V} = \frac{\kappa a}{1 + \kappa a}.$$

Thus the evaluation of $1 - g$ as given by (6.24) could be carried out for dilute solutions of NaCl, fig. 1. The values of x, and hence of the concentration γ (mols per litre), were found from (6.21) and (6.29). At $T = 291°\,K$, $D = 81$, $\beta = 7{\cdot}04 \,.\, 10^{-8}$, $a = 4{\cdot}02 \,.\, 10^{-8}$ cm, $\mu_1 = -2{\cdot}02$. The deviations from experiment at higher concentrations are to be expected as the neglect of the terms in x, $x^{\frac{1}{2}}$, ..., in (6.32) is no longer permissible. According to (6.19), x_0 and hence γ reach a maximum at $t = 2/3$, $\gamma = 0{\cdot}028$, and then begin to decrease with further increase of t (and therefore of F). This anomalous behaviour no longer exists when we use (6.21), as is illustrated by the form (6.36) for dilute solutions.

We shall require the general form of $v(t)$ and $\chi(t)$ and hence of δ and $H(T, V, \epsilon_i)$ at greater concentrations. As will be seen in § 7,

$$H(T, V, \epsilon_i) = H_1(T, \epsilon_i)\, N^2/V + H_2(T, \epsilon_i)\, N^3/V^2 + \ldots, \qquad (6.38)$$

and in place of (6.27),

$$\theta = \delta/\mathrm{D} = \theta_1\gamma^{\frac{1}{2}} + \theta_2\gamma + \theta_3\gamma^{\frac{3}{2}} + \ldots . \tag{6.39}$$

Further, as seen from (6.21) and (6.31), we have for $\chi(t)$,

$$\chi(t) = \chi_2 t^2 + \chi_3 t^3 + \chi_4 t^4 + \ldots \tag{6.40}$$

Also

$$v(t) = v_1 t + v_2 t + v_3 t^3 + \ldots . \tag{6.41}$$

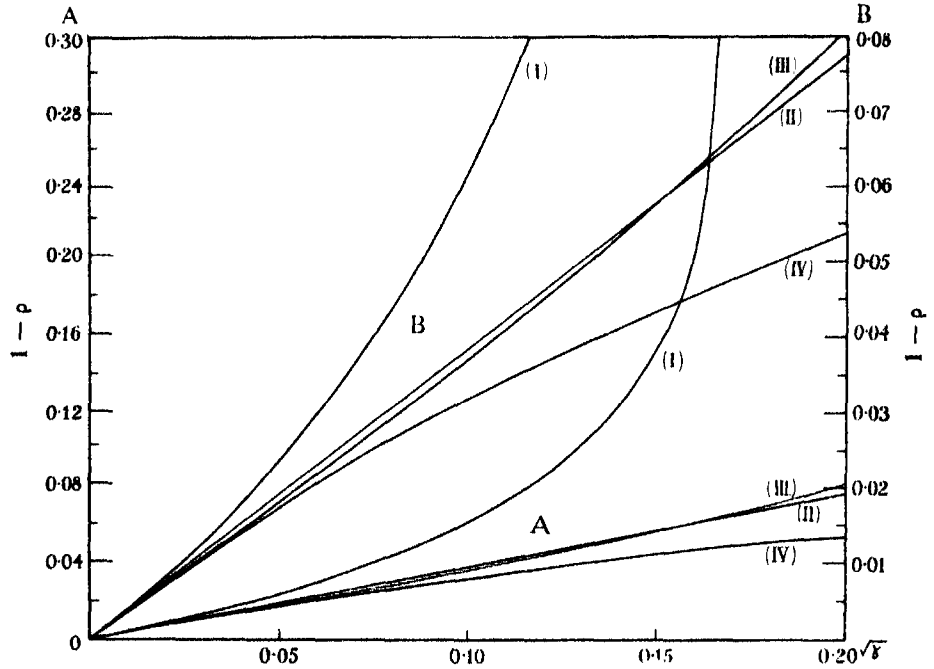


FIG. 1—Variation of osmotic coefficient with concentration for NaCl solution. The group B is a magnification of the group A. In each group the curves are as follows: (i) Curve according to Kramers's theory; (ii) Debye and Hückel limiting curve; (iii) curve for equation (6.24); (iv) experimental curve (given with sufficient accuracy by the Debye and Hückel second approximation).

Here the coefficients θ_1, θ_2, ..., χ_1, χ_2, ..., v_1, v_2, ..., are all functions of $\mathrm{H}_1(\mathrm{T}, \epsilon_i)$, ..., and are calculated as follows. From (6.19) we have that

$$t_0 = \tfrac{2}{3}(\pi x_0)^{\frac{1}{2}} + 2(\pi x_0) + 7(\pi x_0)^{\frac{3}{2}} + \ldots , \tag{6.42}$$

and from (6.21), substituting (6.41) for $v(t)$,

$$t = \tfrac{2}{3}(\pi x)^{\frac{1}{2}} + 2\left(1 - \frac{v_1}{9}\right)(\pi x) + \left(7 - 2v_1 - \tfrac{4}{27}v_2 + \frac{v_1}{9}\right)(\pi x)^{\frac{3}{2}} + \ldots . \tag{6.43}$$

Here * (6.42) is valid for

$$|t_0| < \frac{R^2 |d_1|^2}{6M_0} = 0 \cdot 133, \text{ or for NaCl, } x_0 < 0 \cdot 0076,\ \gamma < 0 \cdot 018,$$

using (6.19); $R = \frac{1}{3}$ is the radius of convergence of

$$x_0^{\frac{1}{2}} = \sum_{r=1}^{\infty} d_r t_0^r,\ d_1 = \tfrac{3}{2}\sqrt{\pi} \text{ and } M_0 = x_0^{\frac{1}{2}}(\tfrac{1}{3}) = 0 \cdot 10,$$

as given by (6.19), is the maximum value of $x_0^{\frac{1}{2}}$ in the given range. Similarly (6.43) is valid for

$$|t| < \frac{R^2 |d_1|^2}{6M} = 0 \cdot 098,$$

where $M = x^{\frac{1}{2}}(\frac{1}{3}) = 0 \cdot 14$, *i.e.*, for NaCl, when $x < 0 \cdot 0041$, $\gamma < 0 \cdot 013$, as obtained from (6.36).

On substituting (6.42) into (6.18), we obtain†

$$\begin{aligned} G(\xi) = \tfrac{1}{2}N \log \omega_0 &= -\tfrac{2}{3}N(\pi x_0)^{\frac{1}{2}} - N(\pi x_0) - \tfrac{7}{3}N(\pi x_0)^{\frac{3}{2}} - \ldots \\ &= -\tfrac{2}{3}AN^{\frac{3}{2}}\xi^{-\frac{3}{2}} - A^2N^2\xi^{-3} - \tfrac{7}{3}A^3N^{\frac{5}{2}}\xi^{-\frac{9}{2}} - \ldots . \end{aligned} \tag{6.44}$$

Similarly, using (6.20), (6.40), and (6.41),

$$\begin{aligned} G_1(\xi_1) = \tfrac{1}{2}N \log \omega &= -\tfrac{2}{3}AN^{\frac{3}{2}}\xi_1^{-\frac{3}{2}} - A^2N^2\xi^{-3}(1 - \tfrac{2}{9}\chi_2 - \tfrac{2}{9}v_1) \\ &- A^3N^{\frac{5}{2}}\xi^{-\frac{9}{2}}\left(\tfrac{7}{3} - \tfrac{4}{3}v_1 - \tfrac{4}{3}\chi_2 + \frac{v_1^2}{9} + \tfrac{4}{27}\chi_2 v_1 - \tfrac{4}{27}\chi_3 - \tfrac{4}{27}v_2\right) - \ldots . \end{aligned} \tag{6.45}$$

Here $G_1(\xi_1)$ is not the same functional form as $G(\xi)$, owing to the fact that ω in (6.16) is not a pure function of x. Hence we write $G_1(\xi_1)$ in place of $G(\xi_1)$ in (6.45).

A first approximation to δ has actually been obtained by using the first terms only in the expansions of $G(\xi)$ and $G_1(\xi_1)$. This gave $\psi(x)$ and hence $v(t)$ and $\chi(t)$. Knowing v_1 and χ_2, we may use the first two terms on the right of (6.44) and (6.45) and substitute in (3.16) to obtain a better approximation to δ and hence to $v(t)$ and $\chi(t)$.

---

* *See* Dienes, " The Taylor Series," p. 260.

† According to Halpern, *loc. cit.*, the free energy F, given by (6.44), may be written as $F = -J(\kappa/3D + C\kappa^2 + \ldots)$ a function of $J = \sum_{i=1}^{s} N_i \varepsilon_i^2$ only (since $\kappa^2 = 4\pi J/DkTV$), and hence must have the form $F = -J\kappa/3D$ in contradiction to (6.44). But we note that $C = J/4(DkT)^2 N$, which Halpern considers to be a constant, involves both J and $N = \sum_{i=1}^{s} N_i$, *i.e.*, F is not a function of J only. Indeed, if J/N = constant, then the form $F = Jf(J)$ used by Halpern reads $F = N\phi(N)$ as assumed in (3.19).

This process may be continued until no longer necessary. At higher concentrations we need to find the analytic continuations of (6.44) and (6.45).

At greater concentrations, equation (6.17) itself becomes unsatisfactory, for it is necessary to include the second term in the series (6.8), so that, in place of (6.9), we have

$$\frac{\partial \bar{A}_N}{\partial \beta} = \frac{N}{\alpha}\frac{\partial \alpha}{\partial \beta}\left[\bar{A}_N - \bar{A}_{N-1} + \frac{N}{1-\theta}\frac{\partial \theta}{\partial N}\bar{A}_{N-2}\right]. \tag{6.46}$$

Noting that

$$\bar{A}_{N-2} = [\omega(x)]^{N-2}\exp\left[-2\frac{N\omega'}{\omega}\frac{\partial x}{\partial N}\right], \tag{6.47}$$

obtained by the same method used in (6.14), (6.46) gives, in place of (6.16),

$$3x\omega' = \omega - \exp\left[-\frac{x\omega'}{\omega}(1+\psi)\right] + \frac{\psi}{3}\exp\left[-\frac{2x\omega'}{\omega}(1+\psi)\right]. \tag{6.48}$$

Substitution of $-t = x\omega'/\omega$ leads to

$$\omega = \frac{e^t e^x}{1+3t}\left(1 + \frac{\psi}{3}e^t e^x\right). \tag{6.49}$$

The corresponding form for x is derived in the same manner as in (6.21), giving

$$\frac{dx}{x} = \frac{(2-3t) - \psi(1+3t)}{t(1+3t)(1+\psi' x)}\,dt + \phi(t)\,dt, \tag{6.50}$$

with solution

$$K^2 x = \frac{t^2}{(1+3t)^3}e^{V(t)}, \quad V(t) = v(t) + \int \phi(t)\,dt. \tag{6.51}$$

Here $v(t)$ is given by (6.22) and

$$\phi(t) = \tfrac{1}{3}\frac{e^t e^x[\psi' + (t' + \chi')\psi]}{t(1+\psi' x)(1+\psi e^t e^x/3)}\frac{dx}{dt}.$$

Successive approximations to $V(t)$ and hence to x and ω may be obtained as before. We shall not proceed further with the solution at this stage of the theory.

7—EVALUATION OF $\bar{\gamma}$ FOR THE FORM $E_{ij} = \lambda_n/(n-1)\,r_{ij}^{\,n-1}$

We shall illustrate the method of finding $\bar{\gamma}$ by choosing a special form for E_{ij} in (2.1). A first approximation to $\bar{\gamma}$, valid for dilute solutions, may be obtained by treating only binary encounters of the ions. Suppose the solution contains two types of ions ε_j and ε_i, in number N_i and N_j respectively, such that $N_i = iN$, $N_j = jN$, *i.e.*, $i + j = 1$. We assume a force

$$F_{ij} = \varepsilon_i\varepsilon_j/Dr_{ij}^2 + \lambda_n/r_{ij}^n, \tag{7.1}$$

so that

$$E_{ij} = \lambda_n/(n-1)\, r_{ij}^{n-1}. \tag{7.2}$$

The suffixes m, n, p shall refer to the pairs (i, i), (i, j), (j, j), respectively.

The contribution to $\bar{\gamma} = \frac{1}{2}\sum_{i,j}' \bar{E}_{ij}$ from binary encounters is given by the sum of three terms corresponding to encounters of the pairs, viz.,*

$$\begin{aligned}\bar{\gamma} &= \sum_{m,n,p} \frac{4\pi N_i N_j}{V} \int_0^\infty \frac{\lambda_n}{n-1} . r_{ij}^{-n+3} \\ &\qquad \times \exp\left[-\frac{1}{kT}\left(\varepsilon_i\varepsilon_j/Dr_{ij} + \lambda_n/(n-1)\, r_{ij}^{n-1}\right)\right] dr_{ij} \\ &= \frac{4\pi N^2}{V} \sum_{m,n,p} \frac{ij\lambda_n}{n-1} \sum_{\tau=0}^{\infty} \frac{(-1)^\tau}{\tau\,!}\left(\frac{1}{kT}\frac{\varepsilon_i\varepsilon_j}{D}\right)^\tau \int_0^\infty r_{ij}^{\tau-n+3} \\ &\qquad \times \exp\left[-\frac{1}{kT}\frac{\lambda_n}{(n-1)\, r_{ij}^{n-1}}\right] dr_{ij} \\ &= \frac{4\pi N^2}{V} \sum_{m,n,p} \frac{ij\lambda_n^2}{(n-1)^2}\left[\frac{\lambda_n}{(n-1)\,kT}\right]^{-(n-4)/(n-1)} \\ &\qquad \times \sum_{\tau=0}^{\infty} \frac{(-1)^\tau}{\tau\,!}\left(\frac{\varepsilon_i\varepsilon_j}{kDT}\right)^\tau \left[\frac{\lambda_n}{(n-1)\,kT}\right]^{-\tau/(n-1)} \Gamma\left(\frac{\tau+n-4}{n-1}\right),\end{aligned} \tag{7.3}$$

i.e.,

$$\bar{\gamma} = \frac{4\pi N^2}{V} \sum_{m,n,p} A_n \sum_{\tau=0}^{\infty} B_{n\tau} T^{\nu'_n - \nu_n \tau},$$

where

$$A_n = ijk^2 \left[\lambda_n/k\,(n-1)\right]^{(n+2)/(n-1)}, \quad \nu'_n = \frac{n-4}{n-1}, \quad \nu_n = \frac{n-2}{n-1},$$

$$B_{n\tau} = \frac{(-1)^\tau}{\tau\,!}\left(\frac{\varepsilon_i\varepsilon_j}{kD}\right)^\tau \left[\frac{\lambda_n}{k\,(n-1)}\right]^{-\tau/(n-1)} \Gamma\left(\frac{\tau+n-4}{n-1}\right).$$

Similar forms may be obtained for $\sum_{i=1}^{N} \varepsilon_i \overline{\frac{\partial\gamma}{\partial\varepsilon_i}}$ and $\overline{\frac{\partial\gamma}{\partial T}}$ provided E_{ij} is a function of ε_i and T. Substituting the resulting expressions for

$$\bar{\gamma},\ \sum_{i=1}^{N} \varepsilon_i \overline{\frac{\partial\gamma}{\partial\varepsilon_i}} \text{ and } \overline{\frac{\partial\gamma}{\partial T}} \text{ into } \chi\,(T, V, \varepsilon_i),$$

---

* *See* Jones, 'Proc. Roy. Soc.,' A, vol. 106, p. 463 (1924).

as given by (3.10), it follows from (3.13) that $H(T, V, \varepsilon_j)$ is of the form $H_1(T, \varepsilon_i) N^2/V$. A treatment of ternary encounters yields a term $H_2(T, \varepsilon_i) N^3/V^2$, and in this way we obtain the series (6.38).

A negative value for $\bar{\gamma}$ implies a tendency for the ions of opposite charge to associate. The (negative) contribution to $\bar{\gamma}$ from (i, j) encounters would be greater than those from the (i, i) and (j, j) interactions,* λ_n/r_{ij}^n being an attractive force. If stronger association exists with solvents of low dielectric constant, $\bar{\gamma}$ and hence δ would become appreciable even for very dilute solutions. This would explain the failure of the Debye-Hückel theory with electrolytes of such solvents and confirm the applicability of the concept of ionic association as was carried out by Fuoss and Kraus.†

If we assume the ions to be rigid spheres, we let $n \to \infty$ subject to the conditions $\sigma_{ij} = \lim\limits_{n \to \infty} \lambda_n^{1/(n-1)}$, where $\sigma_{ij} = \sigma_i + \sigma_j$, σ_i, σ_j, being the radii of the ions i, j, with a similar treatment for m and p. However, in this case $\lim \bar{\gamma} = 0$. This does not in any way invalidate the method developed here, for there certainly exists a term $E_{ij} \neq 0$ due to saturation and hydration effects and to those forces other than the Coulomb between the ions themselves. We shall endeavour to evaluate E_{ij} in a further work.

Finally, there is need of a more satisfactory estimation of the roots b_m in § 4, since the existence of such roots violating the condition (4.15) produces a change in the form (6.3) for the free energy. A more satisfactory proof of the validity of the Gauss distribution law in § 5 is also desirable. The problem of incomplete dissociation may be investigated if E_{ij}, yielding the work of ionization, is known and if the internal partition function of the solute molecules can be calculated. For the equation of mass action would at once lead to the degree of dissociation. We shall return to these points on a later occasion.

APPENDIX

We proceed to examine the legitimacy of Kramers's ensemble. In the usual form of generalized ensemble, the integrand in (3.1) is replaced by the summation‡

$$\sum_{\nu_1, \ldots \nu_s = 0}^{\infty} \exp\left[- E(\nu_i)/kT - \sum_{i=1}^{s} \alpha_i \nu_i\right], \quad E(\nu_i) \equiv E(\nu_1, \ldots, \nu_s), \qquad (A.1)$$

---

* According to Bronsted (*see also* Guggenheim, *loc. cit.*) the contributions from (i, i) and (j, j) pairs would be negligible; ' J. Amer. Chem. Soc.,' vol. 44, p. 877 (1922).

† ' J. Amer. Chem. Soc.,' vol. 55, pp. 476, 1019, 2397 (1933) and later papers.

‡ *Cf.* Pauli, ' Z. Physik,' vol. 41, p. 81 (1927).

subject to the conditions

$$\frac{1}{k\mathrm{T}}\frac{\partial \mathrm{F}}{\partial \mathrm{N}_i} + \alpha_i = 0 \qquad (i = 1, 2, \dots, s). \tag{A.2}$$

There exists an extraordinarily steep maximum in the number of samples in the summation (A.1) at $\nu_i = \mathrm{N}_i$ ($i = 1, 2, \dots, s$), such that only for a variation of ν_i from N_i given by

$$\Delta \mathrm{N}_i/\mathrm{N}_i = |\, \nu_i - \mathrm{N}_i \,|/\mathrm{N}_i \leqslant \mathrm{O}\,(1/\mathrm{N}^{\frac{1}{2}}), \tag{A.3}$$

is the number of samples in (A.1) appreciable. Further, if F_{II} is the free energy obtained when (A.1) is used in (3.1) and F_{I} is that given by (3.1) itself, then the difference between these quantities is given by

$$|\,(\mathrm{F}_{\mathrm{II}} - \mathrm{F}_{\mathrm{I}})/\mathrm{F}_{\mathrm{I}}\,| \sim (\log \mathrm{N})/\mathrm{N}. \tag{A.4}$$

We need to prove that similar properties hold for the distribution of samples employed by Kramers in order that the relation (4.1) be valid. This distribution is governed by the multinomial expansion

$$\begin{aligned}(\mathrm{N}_1 + \dots + \mathrm{N}_s)^{\mathrm{N}} &= \sum_{\nu_1, \dots \nu_s = 0}^{\mathrm{N}} \mathrm{N}\,!\, \prod_{i=1}^{s} (\mathrm{N}_i^{\nu_i}/\nu_i\,!) \\ &= \sum_{\nu_1, \dots \nu_s = 0}^{\mathrm{N}} \mathrm{N}\,!\, \exp\Big[\sum_{i=1}^{s} \nu_i \log \mathrm{N}_i\Big] \Big/ \prod_{i=1}^{s} (\nu_i\,!),\end{aligned} \tag{A.5}$$

where $\sum_{i=1}^{s} \nu_i = \mathrm{N}$, and the term $\mathrm{N}\,!\, \prod_{i=1}^{s} (\mathrm{N}_i^{\nu_i}/\nu_i\,!)$ represents the number of samples which have ν_i ions of kind i, ($i = 1, 2, \dots, s$). Under these conditions (4.1) becomes

$$\begin{aligned}\mathrm{V}^{\mathrm{N}} e^{-\mathrm{F}_{\mathrm{II}}/k\mathrm{T}} &= \int \dots \int \sum_{\nu_1, \dots \nu_s = 0}^{\mathrm{N}} \frac{\mathrm{N}\,!}{\prod_{i=1}^{s} (\nu_i\,!)} \\ &\times \exp\Big[\sum_{i=1}^{s} \nu_i \log \frac{\mathrm{N}_i}{\mathrm{N}} - \mathrm{E}\,(\nu_i)/k\mathrm{T}\Big]\, dx_1\, dy_1\, dz_1 \dots dz_{\mathrm{N}},\end{aligned} \tag{A.6}$$

so that here $\alpha_i = \log \mathrm{N}_i/\mathrm{N}$.

We shall only illustrate with a binary electrolyte, containing $\mathrm{N}_1 + \mathrm{N}_2$ ions. Here (A.5) becomes

$$(\mathrm{N}_1 + \mathrm{N}_2)^{\mathrm{N}} = \sum_{\nu=0}^{\mathrm{N}} \binom{\mathrm{N}}{\nu} \mathrm{N}_1^{\mathrm{N}-\nu} \mathrm{N}_2^{\nu}. \tag{A.7}$$

Putting $\Delta = \nu - \mathrm{N}_2$, the average value of $(\Delta/\mathrm{N})^2$ is

$$\frac{\overline{\Delta^2}}{\mathrm{N}^2} = \frac{1}{\mathrm{N}^{\mathrm{N}}} \sum_{\nu=0}^{\mathrm{N}} \frac{(\nu - \mathrm{N}_2)^2}{\mathrm{N}^2} \binom{\mathrm{N}}{\nu} \mathrm{N}_1^{\mathrm{N}-\nu} \mathrm{N}_2^{\nu} = \frac{\mathrm{N}_2}{\mathrm{N}^2}\Big(1 - \frac{\mathrm{N}_2}{\mathrm{N}}\Big) \sim \frac{1}{\mathrm{N}}. \tag{A.8}$$

Further, by applying Stirling's formula, the ratio of a typical term in (A.7), $N!\,N_1^{N_1-\Delta}N_2^{N_2+\Delta}/(N_1-\Delta)!\,(N_2+\Delta)!$, to the greatest term, $N!\,N_1^{N_1}N_2^{N_2}/N_1!\,N_2!$, is

$$(1+\Delta/N_2)^{-N_2}(1-\Delta/N_1)^{-N_1}(1-\Delta/N_1)^{\Delta}(1+\Delta/N_2)^{-\Delta}$$
$$=1-\frac{N}{2N_1N_2}\Delta^2+O\left(\frac{\Delta^3}{N^2}\right)+\ldots\simeq\exp\left(-\frac{N\Delta^2}{2N_1N_2}\right),\quad\text{(A.9)}$$

if $\Delta\leqslant N^{\frac{2}{3}}$. It is seen from (A.8) and (A.9) that for $\Delta^2>N$ the number of samples decreases very rapidly and thus (A.3) is true. Consideration of more complex electrolytes gives similar results.

For a binary electrolyte (A.6) gives

$$e^{-F_{II}/kT}=\frac{1}{N^N}\sum_{\nu=0}^{N}\binom{N}{\nu}N_1^{N-\nu}N_2^{\nu}e^{-F_I(N_1-\Delta,\,N_2+\Delta)/kT}.\quad\text{(A.10)}$$

Retaining the first three terms in the expansion ($\Delta=\nu-N_2$)

$$F_I(N_1-\Delta,N_2+\Delta)=F_I(N_1,N_2)+(\nu-N_2)\left(\frac{\partial}{\partial N_2}-\frac{\partial}{\partial N_1}\right)F_I$$
$$+\frac{(\nu-N_2)^2}{2}\left(\frac{\partial}{\partial N_2}-\frac{\partial}{\partial N_1}\right)^2F_I+\ldots,\quad\text{(A.11)}$$

we get

$$e^{-(F_{II}-F_I)/kT}=\frac{e^{-\lambda N_2}}{N^N}\sum_{\nu=0}^{N}e^{\lambda\nu}e^{\mu(\nu-N_2)^2}\binom{N}{\nu}N_1^{N-\nu}N_2^{\nu},\quad\text{(A.12)}$$

where

$$\lambda=-\frac{1}{kT}\left(\frac{\partial}{\partial N_2}-\frac{\partial}{\partial N_1}\right)F_I=\log f_1/f_2\sim 1,$$
$$\mu=-\frac{1}{2kT}\left(\frac{\partial}{\partial N_2}-\frac{\partial}{\partial N_1}\right)^2F_I\sim\frac{1}{N}.$$

Neglecting the term in μ, we have

$$e^{-(F_{II}-F_I)/kT}\simeq\frac{e^{-\lambda N_2}}{N^N}\sum_{\nu=0}^{N}e^{\lambda\nu}\binom{N}{\nu}N_1^{N-\nu}N_2^{\nu}$$
$$=\frac{e^{-\lambda N_2}}{N^N}[N_1+N_2e^{\lambda}]^N=\exp\left[N\log\left\{1+\frac{N_2}{N}(e^{\lambda}-1)\right\}-\lambda N_2\right]$$
$$=e^{-\lambda N_2}\left[1+\frac{N_2}{N}(e^{\lambda}-1)\right]^N$$
$$=\exp\left[\frac{N_2\lambda^2}{2}\left(1-\frac{N_2}{N}\right)+\frac{N_2\lambda^3}{6}\left(1-\frac{3N_2}{N}+\frac{2N_2^2}{N^2}\right)+\ldots\right],\quad\text{(A.13)}$$

assuming $|(N_2/N)(e^\lambda - 1)| < 1$, giving

$$|(F_{II} - F_I)/F_I| \sim \frac{N_2}{2N}\left(1 + \frac{N_2}{N}\right)\lambda^2 + \dots . \tag{A.14}$$

If, for the symmetrical type ($N_1 = N_2 = N/2$), we assume $\log f_1 = \log f_2$ then $\lambda = 0$, and this is equivalent to the condition (A.2), so that (A.4) holds. For the non-symmetrical type, however, $\lambda \neq 0$. For example, if for $z_2 = -2z_1$, $N_1 = 2N/3$, $N_2 = N/3$, we use the Debye-Hückel limiting expression for the activity coefficient $\log f_i = -\frac{\varepsilon^2 z_i^2 \kappa}{2D_0 kT}$, we find at $T = 291°$ K, $\lambda = 0{\cdot}59$. Then at $\gamma = 0{\cdot}01$ (A.13) becomes $|(F_{II} - F_I)/F_I| \simeq 0{\cdot}16$,* using the form (3.17) for F_I. Since the greatest contribution to the sum in (A.13) comes from $\nu \sim N$, then the neglected terms in the expansion (A.11) must be considered. Indeed, if F_I is given by (3.17), it is seen that μ is a negative quantity, so that (A.12) will give a smaller value for $|(F_{II} - F_I)/F_I|$ than (A.13).

With symmetrical electrolytes, λ departs from zero as the concentration increases, so that a correction to F_{II} will be necessary. A first approximation to this correction may be obtained by substituting

$$\lambda = -\frac{1}{kT}\left(\frac{\partial}{\partial N_2} - \frac{\partial}{\partial N_1}\right)F_{II}$$

into (A.13). For the high-valency types where it is more probable that $\log f_1 \neq \log f_2$ for dilute solutions and also in the non-symmetrical cases, there will be an appreciable difference between F_I and F_{II}. This is in accordance with the failure of the Debye-Hückel limiting law for such types of electrolytes. The applicability of Kramers's generalized ensemble in these cases thus becomes questionable.

In conclusion, the author wishes to express his sincere thanks for help received from Professor E. F. Burton, Dr. C. Barnes, and Dr. J. K. L. MacDonald of the Department of Physics at the University of Toronto, and from Professor I. A. Barnett of the Department of Mathematics at the University of Cincinnati.

The writer is indebted to the University of Cincinnati for a research fellowship enabling him to continue the work on this paper.

SUMMARY

In this paper, a theory of strong electrolytes is developed by an application of statistical mechanics. The method of Kramers is adopted and

---

* If we use the second approximation of Debye and Hückel, then for K_2SO_4, $a = 2{\cdot}69 \,.\, 10^{-8}$, we obtain $|(F_{II} - F_I)/F_I| \simeq 0{\cdot}13$.

it is shown that if the ordinary Coulomb forces between the ions is used, under certain assumptions, his treatment is valid. Owing to his use of a special generalized ensemble, a correction is necessary which is small for low-valency symmetrical electrolytes but which may be quite considerable for the non-symmetrical cases. The deviations from the inverse square law, due to the saturation and hydration effects on the water dipoles and to the polarization, van der Waals and exchange forces between two typical ions i and j is accounted for by means of a correction term E_{ij} in the expression for their energy of interaction. It is shown that the addition of this term is equivalent to a modification of the dielectric constant D to $D - \delta$, where δ depends on E_{ij}, and is a function of the concentration and temperature. The extension in Kramers's theory as a result of this new type of force is given, and it is seen that with the proper form for δ the method proposed here should satisfactorily describe the properties of electrolytic solutions at strong concentrations. Definite numerical results cannot be obtained until δ is known. The attempt has been made to avoid the essential difficulties in the original Debye-Hückel theory.

The Creation of Electron Pairs by Fast Charged Particles

By H. J. BHABHA, Ph.D., Gonville and Caius College

(*Communicated by R. H. Fowler, F.R.S.—Received April* 29—
Revised June 27, 1935)

INTRODUCTION

We shall discuss in this paper the creation of electron-pairs by the collision of fast charged particles. This calculation goes farther than other calculations on this subject in considering the effect of screening, and in investigating the probability of the creation of a pair as a function of impact parameter, *i.e.*, the least distance of approach between the two colliding particles. We shall also treat certain other cases which have not been considered before, among them the creation of very slow pairs such that the kinetic energies of the electron and positron of the pair are small compared to their rest energy. When the energy of one of the colliding particles is large compared with its rest mass, we shall also show that to a certain approximation most of the formulae given by the direct calculation can be obtained quite simply by a method similar to that given by Weizsäcker for calculating the emission of radiation by fast electrons on colliding with nuclei.

The procedure consists in calculating the probability of the transition of an electron from its initial state of negative energy to a final state of positive energy under the perturbing influence of the two colliding particles, the electron and resulting hole then appearing as the electron and positron of the created air. We shall throughout use the Born approximation, in which the interaction between the particles is treated as a perturbation. The transition from the initial to the final state of the system can then happen in two ways. The electron *in the negative energy state* may either interact with one of the colliding particles and jump at once to its final state, the colliding particle going over into an intermediate state. This particle can then interact with the other colliding particle and both jump to their final states. Or, the electron in the negative energy state may interact with one of the colliding particles and jump to an intermediate state, after which its interaction with the other colliding particle causes it to jump to its final state. Both processes are strictly of the second order, but for brevity we shall call the former process a "first-order process," only in the sense that it involves just one matrix element of the interaction of the electron of the created pair with the colliding particles. The second process involves two matrix elements of the interaction of the electron of the created pair with the colliding particles, and we shall call it a "second-order process."

In the collision between two heavy particles, moving with kinetic energy *small compared with their rest energy*, the "first-order process" is much larger than the second. This is easily seen as follows. We consider the process in a system in which one of the colliding particles is initially at rest, and the other moving with kinetic energy $T = P^2/2M$, where P is its momentum and M its mass. Then its change of momentum ΔP is given by

$$\Delta P \sim \frac{M}{P}\,\Delta T \gtrsim \frac{M}{P}(E + E_+) \gtrsim \frac{Mc}{P}(p + p_+),$$

where E, p, E_+, p_+ are the total energies and momenta of the electron and positron respectively. Since

$$P \ll Mc,$$

$$\Delta P \gg (p + p_+).$$

The total change in momentum of all three particles must be zero, so that the change in momentum of each heavy particle is large compared with the change in momentum of the electron of the created pair. The matrix element for each jump is just the matrix element of the Born approximation for a Coulomb field, and is roughly proportional to $1/(\Delta P)^2$ where one heavy particle goes straight from its initial to its final state, and is proportional to $1/(\mathbf{p} + \mathbf{p}_+)^2$ where the electron jumps straight from its initial to its final state. In the "second-order process," both the heavy particles jump successively from their initial direct to their final states, so that this process is roughly proportional to $1/(\Delta P)^4$. In the "first-order process," the electron and only one of the heavy particles jump *direct* from their initial to their final states, so that this process is proportional to $1/(\Delta P)^2 . (\mathbf{p} + \mathbf{p}_+)^2$, and it follows directly from the inequality $\Delta P \gg p + p_+$ that the "first-order process" is much greater than the second.† For fast particles where $P \gtrsim Mc$, ΔP is no longer much greater than $(p + p_+)$ and the above considerations do not hold.

The "first-order process" depends on the acceleration of the colliding particles, and with increasing energy, the "second-order process" becomes larger than the first. This means that with increasing energy of the colliding particles those processes which depend on the deflection of one of the colliding particles due to its interaction with the other colliding particle eventually become negligible compared with those processes which depend on the interaction of the electron of the created pair with both the colliding particles at once. This is also shown by the results, since the "first-order process" calculated by Heitler and Nordheim for slow heavy particles *decreases* with increasing kinetic energy of the colliding particles, and is more than a thousand times smaller than the "second-order process" calculated by us for fast heavy particles. We may then neglect the interaction between the two colliding particles, and in consequence the deflection of either caused by the other.

If we restrict ourselves to the creation of pairs of total energy small compared with the relative kinetic energy of the colliding particles, we may neglect the deflection of the colliding particles caused by the reaction of the electron of the created pair. Further, one may treat the two colliding particles classically. One may regard one as fixed at the origin and the other as moving along a straight line with uniform velocity $V \simeq c$. The condition under which one may do this is as follows. If δP_1 be the uncertainty in the momentum of the fast particle and δX_1 the uncertainty in its position, then we must have

$$P_1 \gg \delta P_1 > \hbar/\delta X_1,$$

---

† Heitler and Nordheim, 'J. Phys. Rad.,' vol. 5, p. 451 (1934).

so that

$$\delta X_1 \gg \hbar/P_1, \tag{1}$$

and we cannot consider the particle localized in a region smaller than δX_1 given by (1). (For $cP_1 \sim 10^8$ e.v., $\delta X_1 \gg 10^{-13}$ cm.) This distance is much smaller than the order of the distances at which most of the processes of pair creation take place.

We therefore proceed in this paper as follows. The heavier of the two colliding particles, which we shall call the particle 2, and which may be a bare nucleus or a neutral atom of nuclear charge Z_2e, we shall consider as fixed at the origin of co-ordinates. The other colliding particle of charge Z_1e, which we shall call the particle 1, we shall describe as moving classically with uniform velocity V along a straight line, passing the particle 2 at minimum distance b. We now calculate the probability of the transition of an electron from its initial state of negative energy to a state of positive energy under the combined influence of the two colliding particles. To get the total effective cross-section we must integrate this probability over all values of the impact parameter b of the two colliding particles.

This procedure will give a good approximation under the two conditions stated above, firstly that the particle 1 has kinetic energy comparable with or larger than its rest energy, and secondly that the change in energy of this particle, which must be at least $2mc^2$, be small compared with its initial energy.

In the special case in which the particle 1 is an electron, our method implies a neglect of the effect of exchange of this electron with the electron of the created pair. For the region where our calculations are strictly valid, namely, where the energy of the created pair is small compared with the initial energy of the electron 1, it may easily be seen that the effect of exchange is small. Its effect, however, may not be small in the total integrated cross-section, and we shall discuss this point further with the final results.

In § 1 we give the general theory of the pair creation. In § 2 we give the differential and integral cross-sections. The effect of screening will be considered in § 3, and the dependence upon impact parameter investigated in § 4. In § 5 we shall show that to a certain approximation most of the formulae can be derived by a method analogous to that due to v. Weizsäcker for calculating the radiation emitted by fast electrons on colliding with nuclei. The results are discussed in § 6.

1—GENERAL THEORY

We regard the heavier particle, of charge Z_2e, as fixed at the origin of co-ordinates. Its field will be described by a scalar potential $G(r)$. (If it is an unscreened nucleus, this will be Z_2e/r.)

The motion of the other particle, of charge Z_1e, co-ordinates X, Y, Z, will be described by

$$X = b_x; \qquad Y = b_y; \qquad Z = Vt - l, \tag{2}$$

where b is the impact parameter, and l is large compared with atomic dimensions and to b. Its field in the system in which it itself is at rest will be described by a scalar potential $F(r) = Z_1e/r$. In the system we are considering, its field will be described as usual by a scalar and vector potential

$$\left.\begin{aligned} \phi_F &= \frac{1}{\gamma} F\left(\sqrt{(x-b_x)^2 + (y-b_y)^2 + \frac{1}{\gamma^2}(z - Vt + l)^2}\right) \\ A_{Fx} &= A_{Fy} = 0 \\ A_{Fz} &= \frac{V}{c}\phi_F \end{aligned}\right\}, \tag{3}$$

where

$$\gamma = \sqrt{1 - V^2/c^2}.$$

We write the Dirac equation in the form

$$\left\{\frac{E + e\phi}{c} + \left(\boldsymbol{\alpha}, \mathbf{p} + \frac{e}{c}\mathbf{A}\right) + \alpha_4 mc\right\}\psi = 0. \tag{4}$$

We are not interested in the spin of the electron or positron, and we shall have to sum over both directions of spin of the initial, intermediate, and final states. It is here convenient to use a method given by Dirac where this summation is performed automatically.† Our ψ will be a matrix of four rows and columns and a function of x, y, z, t, and hence may be expressed as a linear combination of the α's multiplied by functions of x, y, z, t.

The density will be given by

$$\text{spur}\,[\psi \,.\, \psi^*].$$

We expand ψ in a series

$$\psi = \psi_0 + \psi_1 + \psi_2 + \ldots,$$

---

† Dirac, 'Proc. Camb. Phil. Soc.,' vol. 26, p. 361 (1930).

where ψ_n then satisfies

$$\left\{\frac{E}{c} + (\boldsymbol{\alpha}, \mathbf{p}) + \alpha_4 mc\right\} \psi_n = -\frac{e}{c}\left\{G(r) + \left(1 + \frac{V}{c}\alpha_3\right)\phi_F\right\}\psi_{n-1}. \tag{5}$$

We take the unperturbed initial state ψ_0, which satisfies (5) with the right-hand side put equal to zero, to be of the form

$$\psi_0 = \frac{E_0 - H_0}{c} e^{i\{(\mathbf{p}_0, \mathbf{x}) - E_0 t\}/\hbar}, \tag{6}$$

where

$$H_0 \equiv c(\boldsymbol{\alpha}, \mathbf{p}_0) + \alpha_4 mc^2. \tag{7}$$

ψ_1 will now consist of two parts $\psi_1{}^G$ and $\psi_1{}^F$, corresponding to the two perturbation terms G and ϕ_F in (5). We write

$$\psi_1{}^G = \iiint b^G(\mathbf{p}') \, . \, e^{i\{(\mathbf{p}', \mathbf{x}) - E_0 t\}/\hbar} \frac{d\mathbf{p}'}{h^3}. \tag{8A}$$

Substituting this in (5), multiplying from the left by

$$(E_0 - H')\exp\frac{i}{\hbar}\{-(\mathbf{p}', \mathbf{x}) + E_0 t\}/(E_0{}^2 - H'^2),$$

and integrating over all space ($\mathbf{x}$), we easily find that

$$b^G(\mathbf{p}') = -\frac{(E_0 - H')(E_0 - H_0)}{E_0{}^2 - H'^2}\mathcal{G}(|\mathbf{p}' - \mathbf{p}_0|^2), \tag{8B}$$

where

$$\mathcal{G}(|\mathbf{p}' - \mathbf{p}_0|^2) = \frac{e}{c}\iiint G(r)\, e^{i(\mathbf{p}_0 - \mathbf{p}', \mathbf{x})/\hbar}\, d\mathbf{x}. \tag{9}$$

H' is defined in a similar manner as H_0 in (7) and is a matrix. H'^2 is a number.

ψ_2 will now consist of four terms corresponding to the four terms $-\frac{e}{c}\left\{G(r) + \left(1 + \frac{V}{c}\alpha_3\right)\phi_F\right\}(\psi_1{}^F + \psi_1{}^G)$ on the right-hand side of (5). We are interested only in two of these, $\psi_2{}^{FG}$ and $\psi_2{}^{GF}$, corresponding to the terms $-\frac{e}{c}\left(1 + \frac{V}{c}\alpha_3\right)\phi_F\psi_1{}^G$ and $-\frac{e}{c}G\psi_1{}^F$ in the perturbation. The other two contain the field of one particle only, and cannot lead to the creation of pairs.

We let

$$\psi_2{}^{FG} = \iiiint b^{FG}(\mathbf{p}, E'')\, e^{i\{(\mathbf{p}, x) - E''t\}/\hbar}\frac{d\mathbf{p}\, dE''}{h^4},$$

and get as before by substituting in (5), and using (8),

$$\frac{\mathrm{E}'' + \mathrm{H}}{c}\, b^{\mathrm{FG}}(\mathbf{p}, \mathrm{E}'') = \iiint \frac{\left(1 + \frac{\mathrm{V}}{c}\,\alpha_3\right)(\mathrm{E}_0 - \mathrm{H}')(\mathrm{E}_0 - \mathrm{H}_0)}{\mathrm{E}_0{}^2 - \mathrm{H}'^2} \times \mathscr{G}(|\mathbf{p}' - \mathbf{p}_0|^2)\, \mathrm{F}(\mathbf{p}, \mathrm{E}'', \mathbf{p}', \mathrm{E}_0)\, \frac{d\mathbf{p}'}{h^3}.$$

with

$$\begin{aligned}\mathrm{F}(\mathbf{p}, \mathrm{E}'', \mathbf{p}', \mathrm{E}_0) &= \frac{e}{c\gamma} \iiiint \mathrm{F}\left(\sqrt{(x - \mathrm{X})^2 + (y - \mathrm{Y})^2 + \frac{1}{\gamma^2}(z - \mathrm{Z})^2} \right) \\ &\qquad \times e^{i\{(\mathbf{p}' - \mathbf{p}, \mathbf{x}) - (\mathrm{E}_0 - \mathrm{E}'')t\}/\hbar}\, d\mathbf{x}\, dt \\ &= \frac{e}{c\gamma} \iiint_{-\infty}^{\infty} \mathrm{F}\left(\sqrt{(x - \mathrm{X})^2 + (y - \mathrm{Y})^2 + \frac{1}{\gamma^2}(z - \mathrm{Z})^2} \right) \\ &\qquad \times e^{i(\mathbf{p}' - \mathbf{p}, \mathbf{x} - \mathrm{X})/\hbar}\, d(\mathbf{x} - \mathrm{X}) \,.\int_{-\infty}^{\infty} e^{i\{(\mathbf{p}' - \mathbf{p}, \mathrm{X}) - (\mathrm{E}_0 - \mathrm{E}'')t\}/\hbar}\, dt \\ &= \mathscr{F}\{(p_x - p'_x)^2 + (p_y - p'_y)^2 + \gamma^2(p_z - p'_z)^2\} \\ &\qquad \times h\delta\{\mathrm{V}(p'_z - p_z) - \mathrm{E}_0 + \mathrm{E}''\}\, e^{i/\hbar(\mathbf{p}' - \mathbf{p}, \mathbf{x}_0)},\end{aligned}$$

using (2). Here

$$\begin{aligned}\mathscr{F}_\gamma(|\mathbf{p} - \mathbf{p}'|^2) &\equiv \mathscr{F}\{(p_x - p'_x)^2 + (p'_y - p_y)^2 + \gamma^2(p'_z - p_z)^2\} \\ &= \frac{e}{c} \iiint \mathrm{F}(r)\, e^{i\{(p'_x - p_x)x + (p'_y - p_y)y + \gamma(p'_z - p_z)z\}/\hbar}\, d\mathbf{x}, \qquad (10)\end{aligned}$$

and $\mathbf{X}_0$ denotes the value of $\mathbf{X}$ at $t = 0$ $(b_x, b_y, -l)$. Owing to the presence of the δ function in $\mathrm{F}(\mathbf{p}, \mathrm{E}'', \mathbf{p}', \mathrm{E}_0)$ the integration with respect to p'_z may be carried out, and we obtain

$$\begin{aligned}b^{\mathrm{FG}}(\mathbf{p}, \mathrm{E}'') &= \frac{c}{\mathrm{V}} \iint_{-\infty}^{\infty} \frac{(\mathrm{E}'' - \mathrm{H})\left(1 + \frac{\mathrm{V}}{c}\,\alpha_3\right)(\mathrm{E}_0 - \mathrm{H}')(\mathrm{E}_0 - \mathrm{H}_0)}{(\mathrm{E}''^2 - \mathrm{H}^2)(\mathrm{E}_0{}^2 - \mathrm{H}'^2)} \\ &\qquad \times \mathscr{F}_\gamma(|\mathbf{p} - \mathbf{p}'|^2)\, \mathscr{G}(|\mathbf{p}' - \mathbf{p}_0|^2)\, e^{i(\mathbf{p}' - \mathbf{p}, \mathrm{X}_0)/\hbar}\, dp'_x\, dp'_y, \qquad (11)\end{aligned}$$

taken at the point $p'_z = \mathrm{P}_0 \equiv -\left(\frac{\mathrm{E}'' - \mathrm{E}_0}{\mathrm{V}} - p_z\right)$. We notice that b^{FG} has two singularities at $\mathrm{E}''^2 = \mathrm{H}^2$. In $\psi_2{}^{\mathrm{FG}}$ the E'' integration extends from $-\infty$ to ∞. We deflect the path of integration above the singularities, and for large t the integral then reduces to the residues at these points. Only the one with E'' positive is of interest to us here.

We finally get

$$\psi_2{}^{\mathrm{FG}} = \frac{1}{\mathrm{E}} \iiint \mathrm{J} e^{i\{(\mathbf{p}, \mathbf{x}) - \mathrm{E}t\}/\hbar}\, \frac{d\mathbf{p}}{h^3},$$

with

$$\mathrm{J} = -\frac{\pi i}{h^3}\frac{c}{\mathrm{V}}(\mathrm{E}-\mathrm{H})\left(1+\frac{\mathrm{V}}{c}\alpha_3\right)\{\mathrm{L}_0-(\boldsymbol{\alpha},\mathbf{L})-\alpha_4\mathrm{L}_4\}(\mathrm{E}_0-\mathrm{H}_0), \tag{11}$$

where

$$(\mathrm{L}_0, \mathrm{L}_1 \ldots \mathrm{L}_4) \equiv -\frac{1}{c^2}\iint_{-\infty}^{\infty}(\mathrm{E}_0, cp'_x, cp'_y, c\mathrm{P}_0, mc^2)$$
$$\times \mathscr{J}\{(p_x-p'_x)^2+(p_y-p'_y)^2+\gamma^2(p_z-\mathrm{P}_0)^2\}$$
$$\cdot\frac{\mathscr{G}(|\mathbf{p}'-\mathbf{p}_0|^2)}{p'^2_x+p'^2_y+\mathrm{P}_0^2-p_0^2}\,e^{i(\mathbf{p}'-\mathbf{p}_0,\mathbf{X}_0)/\hbar}\,dp'_x\,dp'_y,$$

and our notation means that E_0 is to be taken in the integral for L_0, cp'_x for L_1, etc. Writing $\mathbf{p}'_r$ for $\mathbf{p}'_r-\mathbf{p}_r$, we get

$$(\mathrm{L}_0, \mathrm{L}_1, \ldots, \mathrm{L}_4) = -\frac{1}{c^2}\iint_{-\infty}^{\infty}\{\mathrm{E}_0, c(p'_x+p_x), c(p'_y+p_y), c\mathrm{P}_0, mc^2\}$$
$$\times\frac{\mathscr{G}\{(\mathbf{p}'_r-\mathbf{p}_{0r}+\mathbf{p}_r)^2+\delta^2\}\,\mathscr{J}(\mathbf{p}'^2_r+\varepsilon^2)}{(\mathbf{p}'_r+\mathbf{p}_r)^2+\mathrm{P}_0^2-p_0^2}\,e^{i(\mathbf{p}'_r,\mathbf{b}_r)/\hbar}\,dp'_x\,dp'_y\,e^{i\epsilon l/\hbar\gamma}. \tag{12}$$

(For definition of δ, ε, *see* (15).) The suffix r will be used to denote a two-vector with components along the x and y axes.

We similarly find for ψ_2^{GF}

$$\psi_2^{\mathrm{GF}} = \frac{1}{\mathrm{E}}\iiint \mathrm{I}e^{i\{(\mathbf{p},\mathbf{x})-\mathrm{E}t\}/\hbar}\frac{d\mathbf{p}}{h^3},$$

with

$$\mathrm{I} = -\frac{\pi i}{h^3}\frac{c}{\mathrm{V}}(\mathrm{E}-\mathrm{H})\{\mathrm{K}_0-(\boldsymbol{\alpha},\mathbf{K})-\alpha_4\mathrm{K}_4\}\left(1+\frac{\mathrm{V}}{c}\alpha_3\right)(\mathrm{E}_0-\mathrm{H}_0), \tag{13}$$

and

$$(\mathrm{K}_0, \mathrm{K}_1, \mathrm{K}_2, \mathrm{K}_3, \mathrm{K}_4) = -\frac{1}{c^2}\iint_{-\infty}^{\infty}\{\mathrm{E}, -c(p'_x-p_{0x}), -c(p'_y-p_{0y}), c\mathrm{P}, mc^2\}$$
$$\cdot\frac{\mathscr{G}\{(\mathbf{p}'_r-\mathbf{p}_{0r}+\mathbf{p}_r)^2+\delta^2\}\,\mathscr{J}(\mathbf{p}'^2_r+\varepsilon^2)}{(\mathbf{p}'_r-\mathbf{p}_{0r})^2+\mathrm{P}^2-p^2} \tag{14}$$
$$e^{i(\mathbf{p}'_r,\mathbf{b}_r)/\hbar}\,dp'_x\,dp'_y\,e^{i\epsilon l/\hbar\gamma}.$$

$$\left.\begin{aligned}
\mathrm{P}_0 = -\left(\frac{\mathrm{E}-\mathrm{E}_0}{\mathrm{V}}-p_z\right); \quad \mathrm{P} &= \frac{\mathrm{E}-\mathrm{E}_0}{\mathrm{V}}+p_{0z}\\
|\mathrm{P}-p_z| = |\mathrm{P}_0-p_{0z}| &= \frac{\mathrm{E}-\mathrm{E}_0}{\mathrm{V}}+p_{0z}-p_z \equiv \delta\\
\gamma|\mathrm{P}-p_{0z}| = \gamma|\mathrm{P}_0-p_z| &= \gamma\,\frac{\mathrm{E}-\mathrm{E}_0}{\mathrm{V}} \equiv \varepsilon\\
\mathrm{P}^2-p^2 &= \left(\frac{\mathrm{E}-\mathrm{E}_0}{\mathrm{V}}+p_{0z}-p\right)\left(\frac{\mathrm{E}-\mathrm{E}_0}{\mathrm{V}}+p_{0z}+p\right)\\
\mathrm{P}_0^2-p_0^2 &= \left(\frac{\mathrm{E}-\mathrm{E}_0}{\mathrm{V}}-p_z-p_0\right)\left(\frac{\mathrm{E}-\mathrm{E}_0}{\mathrm{V}}-p_z+p_0\right)
\end{aligned}\right\}. \tag{15}$$

The initial state ψ_0 given by (6) corresponds to an electron density of $8E_0^2/c^2$ instead of 2, so that we must divide the density given by spur $[\psi_2 \cdot \psi_2^*]$ by $4E_0^2/c^2$ to get the final probability. We thus find that the probability of the creation of a pair, the electron of which has a momentum lying in an element $d\mathbf{p}$ and the positron a momentum lying in an element $d(-\mathbf{p}_0)$, is

$$\frac{c^2}{4E^2E_0^2}\,\text{spur}\,(I+J)(I^*+J^*)\,\frac{d\mathbf{p}\,d\mathbf{p}_0}{h^6}\,. \tag{16}$$

I, J are given by (11) and (13) and I + J may be written in the form

$$I+J = -\frac{\pi i}{h^3}\frac{c}{V}(E-H)\{M_0-(\alpha,\mathbf{M})-\alpha_4 M_4-(\sigma,\mathbf{N}) - \alpha_4\alpha_3 N_4\}(E_0-H_0), \tag{17}$$

where σ represents the matrices $\alpha_2\alpha_3$, $\alpha_3\alpha_1$, $\alpha_1\alpha_2$, and the M's and N's are given by

$$M_0, M_1, M_2, M_3, \dots, N_4 = -\frac{1}{c^2}\iint_{-\infty}^{\infty} \mathcal{J}\{(\mathbf{p}'_r-\mathbf{p}_{0r}+\mathbf{p}_r)^2+\delta^2\}\,\mathcal{J}\{p'^2_r+\varepsilon^2\} \times [M'_0, M'_1, \dots, N'_4]\,e^{i(\mathbf{p}'_r\cdot\mathbf{b}_r)/\hbar}\,dp'_x\,dp'_y\,e^{i/\hbar\, d/\gamma}, \tag{18}$$

with

$$\left.\begin{aligned}
M'_0 &= \frac{E_0-Vp_{0z}}{D}+\frac{E-Vp_z}{D_0}\\
M'_1 &= -\frac{c(p'_x-p'_{0x})}{D}+\frac{c(p'_x+p_x)}{D_0}\\
M'_2 &= -\frac{c(p'_y-p_{0y})}{D}+\frac{c(p'_y+p_y)}{D_0}\\
M'_3 &= \frac{c}{V}\left\{\frac{\gamma^2E-E_0+Vp_{0z}}{D}+\frac{\gamma^2E_0-E+Vp_z}{D_0}\right\}\\
M'_4 &= mc^2\left(\frac{1}{D}+\frac{1}{D_0}\right)\\
N'_1 &= -\frac{V}{c}\left\{\frac{c(p'_y-p_{0y})}{D}+\frac{c(p'_y+p_y)}{D_0}\right\}\\
N'_2 &= \frac{V}{c}\left\{\frac{c(p'_x-p_{0x})}{D}+\frac{c(p'_x+p_x)}{D_0}\right\}\\
N'_3 &= 0;\quad N'_4 = \frac{V}{c}mc^2\left(\frac{1}{D}-\frac{1}{D_0}\right)
\end{aligned}\right\}, \tag{19}$$

and

$$\left.\begin{aligned}
D &= (\mathbf{p}'_r-\mathbf{p}_{0r})^2+P^2-p^2\\
D_0 &= (\mathbf{p}'_r+\mathbf{p}_r)^2+P_0^2-p_0^2
\end{aligned}\right\}. \tag{20}$$

The spur in (16) may then be evaluated by the usual methods if one remembers that the spur of all the Dirac matrices is zero, and also of their products, excepting those giving the unit matrix, whose spur is 4. The final probability (16) may then be written in the form

$$\pi^2 \frac{c^4}{V^2} S \frac{d\mathbf{p}\, d\mathbf{p}_0}{h^{12}}, \tag{21}$$

with

$$\begin{aligned} S = \frac{4}{EE_0}\Big[& |M_0|^2 \{EE_0 + c^2(\mathbf{p}, \mathbf{p}_0) + m^2c^4\} + |\mathbf{M}|^2 \{EE_0 - c^2(\mathbf{p}, \mathbf{p}_0) - m^2c^4\} \\ & + \{|M_4|^2 + |\mathbf{N}|^2\}\{EE_0 - c^2(\mathbf{p}, \mathbf{p}_0) + m^2c^4\} \\ & + |N_4|^2 \{EE_0 + c^2(\mathbf{p}, \mathbf{p}_0) - 2c^2 p_z p_{0z} - m^2c^4\} \\ & + \{Emc^2(M_0M_4^* + M_3N_4^*) + E_0mc^2(M_0M_4^* - M_3N_4^*) \\ & + c(E\mathbf{p}_0 + E_0 p, \mathbf{M})M_0^* + c(E\mathbf{p}_0 - E_0\mathbf{p}, \mathbf{M} \times \mathbf{N}^*) \\ & - c(Ep_{0z} - E_0p_z)M_4N_4^* + c^2(\mathbf{p}, \mathbf{M})(\mathbf{p}_0, \mathbf{M}^*) \\ & + c^2(\mathbf{p}, \mathbf{N})(\mathbf{p}_0, \mathbf{N}^*) + c^2(\mathbf{p}_0 \times \mathbf{p}, \mathbf{N})M^*_0 \\ & + c(p_z - p_{0z})M_0N_4^*mc^2 + c(\mathbf{p}_0 + \mathbf{p}, \mathbf{M})M_4^*mc^2 \\ & - c(\mathbf{N} \times [\mathbf{p} + \mathbf{p}_0])_z N_4^*mc^2\} + \text{conj. complex}\Big]. \end{aligned} \tag{22}$$

We see at once that this expression is quite symmetrical in $\mathbf{p}$ and $\mathbf{p}_0$ if we remember that the M's go over into M* and the N's into − N* on interchanging $\mathbf{p}$ and $\mathbf{p}_0$.

(21) gives the probability of the creation of a pair as function of the impact parameter $\mathbf{b}_r$. To get the effective cross-section for the production of such a pair, we must integrate over all $\mathbf{b}_r$. This is easily performed as follows. $\mathbf{b}_r$ only occurs in the M's and N's, and in S we have only products of two of these. Then, for example,

$$\begin{aligned} \iint_{-\infty}^{\infty} M_\alpha M^*_\beta\, db_x\, db_y = \iint_{-\infty}^{\infty} & \Big\{ -\frac{1}{c^2} \iint \mathcal{G}\{(\mathbf{p}'_r - \mathbf{p}_{0r} + \mathbf{p}_r)^2 + \delta^2\} \\ & \qquad \times \mathcal{K}(p'^2 + \varepsilon^2) M'_\alpha(\mathbf{p}'_r)\, e^{i(\mathbf{p}'_r, \mathbf{b}_r)/\hbar}\, dp'_x\, dp'_y \Big\} \\ & \cdot \Big\{ -\frac{1}{c^2} \iint_{-\infty}^{\infty} \mathcal{G}\{(\mathbf{p}''_r - \mathbf{p}_{0r} + \mathbf{p}_r)^2 + \delta^2\} \\ & \qquad \times \mathcal{K}(p''^2_r + \varepsilon^2) M'_\beta(\mathbf{p}''_r)\, e^{-i(\mathbf{p}r'', \mathbf{b}r)/\hbar}\, dp''_x\, dp''_z \Big\}\, db_x\, db_y. \end{aligned} \tag{23}$$

We interchange the order of the integrations and carry out the $\mathbf{b}_r$ integration first. This immediately gives us two δ functions, $h^2\,\delta(p'_x - p''_x)$

$\delta(p'_y - p''_y)$, so that the integrations with respect to p_x'', p''_y may be performed, and we are finally left with

$$\frac{h^2}{c^4}\iint_{-\infty}^{\infty} |\mathcal{G}\{(\mathbf{p}'_r - \mathbf{p}_{0r} + \mathbf{p}_r)^2 + \delta^2\}|^2 \\ \times |\mathcal{F}\{\mathbf{p}'^2_r + \varepsilon^2\}|^2 \, \mathrm{M}'_\alpha(\mathbf{p}'_r)\, \mathrm{M}'_\beta(\mathbf{p}_r')\, dp'_x\, dp'_y . \quad (23\text{A})$$

Thus the differential effective cross-section $d\mathrm{Q}$ for the creation of a pair, the electron of which has a momentum between $\mathbf{p}$ and $\mathbf{p} + d\mathbf{p}$, and the positron a momentum between $\mathbf{p}_+$ and $\mathbf{p}_+ + d\mathbf{p}_+$, may be written

$$d\mathrm{Q} = \frac{\pi^2}{\mathrm{V}^2}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{S}' \, |\mathcal{G}\{(\mathbf{p}'_r - \mathbf{p}_{0r} + \mathbf{p}_r)^2 + \delta^2\}|^2 \\ \times |\mathcal{F}\{p'^2_r + \varepsilon^2\}|^2 \, dp'_x\, dp'_y\right] \frac{d\mathbf{p}\, d\mathbf{p}_0}{h^{10}}, \quad (24)$$

where S′ is the same expression as (22) with the M's and N's replaced by the M″s and N″s respectively

When both colliding particles with charges Z_1 and Z_2 are unscreened, we get for $\mathcal{F}$ and $\mathcal{G}$

$$\left.\begin{aligned} \mathcal{F}(|\mathbf{p} - \mathbf{p}'|^2) &= -\frac{\mathrm{Z}_1 e^2 h^2}{c\pi}\frac{1}{|\mathbf{p} - \mathbf{p}'|^2} \\ \mathcal{G}(|\mathbf{p} - \mathbf{p}'|^2) &= -\frac{\mathrm{Z}_2 e^2 h^2}{c\pi}\frac{1}{|\mathbf{p} - \mathbf{p}'|^2} \end{aligned}\right\} . \quad (25)$$

Screening will be considered in § 5.

2—THE DIFFERENTIAL AND INTEGRAL CROSS-SECTIONS

The integrals (24) can only be evaluated approximately in certain cases. The rather complicated calculations have been performed in another paper.† These calculations involve no further physical assumptions. We shall give only the results here, together with the conditions which lie at the bottom of the mathematical approximations.

1. We first consider the case where the electron and positron of the created pair have momenta small compared to mc, so that we may neglect terms of the order p^2, p_+^2 compared with m^2c^2

$$p, p_+ \ll mc. \quad (26)$$

We henceforth write $\mathrm{E}_+ = -\mathrm{E}_0$ for the energy of the positron, and $\mathbf{p}_+ = -\mathbf{p}_0$ for its momentum. The differential cross-section for the creation of a pair, the electron of which has a momentum between p and

---

† 'Proc. Camb. Phil. Soc.,' vol. 31, p. 394 (1935). Referred to here as B.

$p + dp$, lying in a small solid angle $d\omega$, and the positron a momentum between p_+ and $p_+ + dp_+$, lying in the solid angle $d\omega_+$, is then (B, § 1, formulae (12) and (13)).

$$dQ = \frac{1}{128\pi^3}\left(\frac{Z_1 Z_2}{137}\right)^2\left(\frac{e^2}{mc^2}\right)^2\left[(p_{+r}^2 + p_r^2)\left\{\log\frac{1}{\gamma^2} - \frac{17}{6} + C\right\}\right.$$
$$\left. + (\mathbf{p}_{+r} + \mathbf{p}_r)^2\left\{\frac{1}{10} + C_r\right\} + (p_{+z} + p_{+z})^2\left\{\frac{1}{20} + C_z\right\}\right]$$
$$\times\frac{p^2 p_+^2\, dp\, dp_+\, d\omega\, d\omega_+}{(mc)^8}, \quad (27)$$

where

$$\left.\begin{aligned} C &= 4\frac{\gamma^2}{1-\gamma^2}\log\frac{1}{\gamma^2} - \frac{4}{3}\gamma^2 + \frac{1}{6}\gamma^4 \\ C_z &= 3\frac{\gamma^2}{1-\gamma^2}\left(1 - \frac{\gamma^2}{1-\gamma^2}\log\frac{1}{\gamma^2}\right) - \frac{13}{5}\gamma^2 + \frac{7}{4}\gamma^4 - \frac{9}{10}\gamma^6 + \frac{1}{5}\gamma^8 \\ C_r &= -\frac{3}{2}\frac{\gamma^2}{1-\gamma^2}\left(1 - \frac{\gamma^2}{1-\gamma^2}\log\frac{1}{\gamma^2}\right) + \frac{4}{5}\gamma^2 - \frac{1}{8}\gamma^4 - \frac{1}{20}\gamma^6 + \frac{1}{40}\gamma^8 \end{aligned}\right\}. \quad (28)$$

This cross-section is accurate for all values of γ. The C's tend to zero as the energy $(1/\gamma)$ of the particle creating the pair increases, so that for large energies we may neglect them in (28). For $1/\gamma \gg 10$, the C's are already less than one-tenth of the corresponding terms in curly brackets in (28). It is also easily shown that each expression in curly brackets in (28), and hence dQ, tends to zero as $\gamma \rightarrow 1$, *i.e.*, as $V \rightarrow 0$. By a numerical calculation we find that the coefficients of $(p_r^2 + p_{+r}^2)$, $(\mathbf{p}_r + \mathbf{p}_{+r})^2$, and $(p_z + p_{+z})^2$ in curly brackets in (28) are 0·08, 0·02, and 0·03 for $1/\gamma = 2$, 0·87, 0·08, and 0·06 for $1/\gamma = 5$, and 1·95, 0·09, and 0·05 for $1/\gamma = 10$.

If we write E_K for the total kinetic energy of the pair, so that

$$E_K = \frac{1}{2m}(p_+^2 + p^2), \quad (29)$$

then the integration of (27) gives the cross-section for the creation of a pair of kinetic energy E_K and total energy $2mc^2 + E_K$, namely,

$$\frac{1}{32}\left(\frac{Z_1 Z_2}{137}\right)\left(\frac{e^2}{mc^2}\right)^2\left\{\log\frac{1}{\gamma^2} - \frac{161}{60} + C + C_r + C_z\right\}\frac{E_K^3\, dE_K}{m^4 C^8}. \quad (30)$$

2. We next consider the case where

$$mc^2 \ll E, E_+;\ mc^2/\gamma \gg E, E_+, \quad (31)$$

the fulfilment of which necessarily requires that $\gamma \ll 1$. In accordance with (31), we neglect terms of order $\gamma^2 E^2$, $\gamma^2 E_0^2$, $(m^2c^4/E)^2$ and $(m^2c^4/E_0)^2$

compared with m^2c^4. The differential cross-section for the creation of a pair, the electron of which has an energy between E and E + dE, and the positron an energy between E_+ and $E_+ + dE_+$, is now (B, § 2 (23))

$$dQ = \frac{8}{\pi}\left(\frac{Z_1Z_2}{137}\right)^2\left(\frac{e^2}{mc^2}\right)^2\frac{E_+{}^2 + E^2 + \frac{2}{3}EE_+}{(E + E)^4}\log\frac{k\,EE_+}{(E + E_+)\,mc^2} \times \log\frac{k'mc^2}{\gamma\,(E + E_+)}\,dE\,dE_+. \quad (32)$$

We shall use k, k' to denote numbers of the order unity, but not always the same numbers. This formula disagrees with that given by Fury and Carlson in having an extra logarithmic term. It has also been obtained recently by Landau and Lifschitz† by a similar method to that given here.

If we write E_T for the total energy of the created pair, *i.e.*,

$$E_T = E + E_+, \quad (33)$$

then the cross-section for the creation of a pair whose total energy lies between E_T and $E_T + dE_T$ is

$$dQ = \frac{56}{9\pi}\left(\frac{Z_1Z_2}{137}\right)^2\left(\frac{e^2}{mc^2}\right)^2\log\frac{kE_T}{mc^2}\log\frac{k'mc^2}{\gamma E_T}\frac{dE_T}{E_T}, \quad (34)$$

subject to the condition $E_T \gg mc^2$. In deriving (34) we have integrated (32) over all E from mc^2 to E_T, whereas (32) is accurate only subject to the condition (31). One may show, however, that the accurate formula in the region $E \sim mc^2$ gives a cross-section smaller than that given by (32), so that the error we make in carrying the E integration down to mc^2 is small if $E_T \gg mc^2$.

To get the total cross-section, we notice that the integral of (34) with respect to E_T varies as $\log^2 E_T$, so that it is not very sensitive to the limits of integration. We may then, without appreciable error, carry the integration from $2mc^2$ to mc^2/γ, and get

$$Q = \frac{28}{27\pi}\left(\frac{Z_1Z_2}{137}\right)^2\left(\frac{e^2}{mc^2}\right)^2\log^3\frac{k}{\gamma}. \quad (35)$$

The formula (35) will give the total cross-section accurately provided that in those regions where (34) fails, namely, when $E_T \sim 2mc^2$ and $E_T \sim mc^2/\gamma$, the accurate formulae do not give cross-sections larger than (34) by an order of magnitude. It can now be shown, provided that exchange effects play no part, which is certainly true if the particle 1 be not an electron, that the accurate formulae in every case given results smaller than (34) in those regions where (34) is inaccurate (one can see this for the region $E_T \sim 2mc^2$ by a comparison of (34) and (30)). It is therefore justifiable to integrate (34) from $E_T \simeq 2mc^2$ to $E_T \simeq mc^2/\gamma$ when $\gamma \ll 1$. We shall discuss the effects of exchange in § 6.

---

† 'Phys. Z. Sowjet.,' vol. 6, p. 244 (1934). The formula (32) is the only result of this paper derived by Landau and Lifschitz.

In deriving (32), (34), and (35) we have neglected the effect of screening of the nucleus 2. It will be seen in § 3 that this is legitimate when

$$\frac{EE_+}{(E+E_+)\,mc^2} \ll \frac{137}{2} Z_2^{-\frac{1}{3}}. \tag{36}$$

The formula (32) is therefore accurate, subject to (36). It will appear in the next section that the effect of screening is only to change the expression under the first logarithm in (32) so that the deviations from (32) when (36) is not fulfilled are not very large. For lead, $137\,Z_2^{-\frac{1}{3}} \sim 31$, so that from (36) we must have $E \sim E_+ \ll 31\,mc^2$. Since (32) is further subject to the condition (31), the region of validity of (32) is small. In the next section we shall give the formulae which are not subject to the restriction (36).

For a given $E_T = E + E_+$, the left-hand side of (36) is greatest when $E = E_+$, and a large contribution to (34) comes from just these regions where $E \sim E_+$, so that the accuracy of (34) is subject to the condition $E_T \ll 2\,.\,137\,Z_2^{-\frac{1}{3}}$. Hence (35) will represent a fairly good approximation, provided

$$1/\gamma < 2\,.\,137\,Z_2^{-\frac{1}{3}}. \tag{44}$$

3. When the particle 1 is a heavy particle with rest mass large compared with that of the electron, a further case is of interest, namely,

$$mc^2/\gamma \ll E, E_+ ;\quad M_1c^2/\gamma \gg E, E_+ . \tag{37}$$

We now neglect terms of order mc^2 compared with γE, γE_0, and consider γ^2 small compared with unity. The differential cross-section for the creation of a pair, the electron of which has an energy E, and the positron an energy E_+ is then (B, § 3 (33))

$$dQ = \frac{8}{\pi}\left(\frac{Z_1Z_2}{137}\right)^2\left(\frac{e^2}{mc^2}\right)^2\frac{m^2c^4}{(E+E_+)^4\,\gamma^2}\log\frac{2k}{\gamma}\,dE\,dE_+, \tag{38}$$

subject to the further condition $\gamma \ll 1$ and $|E - E_+| \ll (E + E_+)$. As regards the term γ^2 in the denominator of (38), we note that (38) is valid only when $E\gamma$, $E_+\gamma \gg mc^2$. If we write $E = E_+ = mc^2/\gamma$ in (32) and (38), *i.e.*, extrapolate both formulae into regions where the conditions for their validity (31) and (37) respectively are not fulfilled, we find that the two formulae go over into each other except that (32) is larger than (38) by a factor 8/3. This means that when (38) is extrapolated down to the region $E \sim E_+ \sim mc^2/\gamma$ it is not in error by more than a factor three.

To get the total cross-section for the creation of a pair with $E, E_+ > mc^2/\gamma$, we integrate (38) over all E, E_+ from mc^2/γ to M_1c^2/γ. We get

$$Q_3 \sim \frac{1}{3\pi}\left(\frac{Z_1Z_2}{137}\right)^2\left(\frac{e^2}{mc^2}\right)^2 \log\frac{2k}{\gamma}. \tag{39}$$

The largest contribution to (39) comes from the region $E, E_+ \sim mc^2/\gamma$, so that from what has already been said above regarding the error in extrapolating (38), (39) represents no more than the correct order of magnitude. It should not be in error by more than a factor three. We notice that Q_3 increases with the logarithm of the energy of the particle 1.

3—THE EFFECT OF SCREENING

We shall now suppose that the fixed nucleus of charge Z_2 is surrounded by a distribution of electrons so as to form a neutral atom. The screening effect of this distribution of electrons may easily be taken into account if we write for the matrix element $\mathcal{G}(q^2)$ instead of (25)

$$G(\mathbf{q}^2) = -\frac{Z_2e^2h^2}{c\pi}\cdot\frac{1 - F(\mathbf{q}^2)}{|\mathbf{q}|^2} \tag{40}$$

by a usual transformation. $F(q^2)$ is the atomic form factor defined by

$$F(\mathbf{q}^2) = \frac{1}{Z_2}\int \rho e^{i(\mathbf{q}\cdot\mathbf{x})/\hbar}\,d\mathbf{x} \tag{41}$$

integrated over the whole of space. ρ is the density of electrons at a distance r from the nucleus. We shall assume that the electron density is that given by the statistical method of Fermi. Now ρ is considerable only inside a region of about $137\,Z_2^{-1/3}\hbar/mc$, which is roughly the radius of the Fermi atom, so that by (41) $F(q^2)$ will be small compared with unity if

$$q \gg \delta' \equiv \frac{\hbar}{137\,Z_2^{-1/3}\hbar/mc} = \frac{Z_2^{1/3}}{137}\cdot mc\,. \tag{42}$$

By (18) $q^2 = \{(\mathbf{p}'_r - \mathbf{p}_{0r} + \mathbf{p}_r)^2 + \delta^2\}$, so that the smallest value of q that occurs is δ. Screening is therefore negligible if

$$\delta \gg \delta' \equiv \frac{Z_2^{1/3}}{137}\,mc, \tag{43}$$

in which case (40) goes over into (25).

1. In the creation of slow pairs, given by (27) $\delta \gtrsim 2mc$ by (15), so that screening has no effect even for the heaviest atoms.

2. For the case of fast pairs given by (32), the particles of the pair are ejected within small angles of the order mc^2/E, mc^2/E_+ about the Z axis, so that by (15)

$$\delta \sim \frac{mc}{2}\frac{(\mathrm{E}+\mathrm{E}_+)\,mc^2}{\mathrm{E}\,\mathrm{E}_+}.$$

Screening is therefore negligible if

$$\frac{\mathrm{EE}_+}{(\mathrm{E}+\mathrm{E}_+)} \ll \frac{137}{2}\,\mathrm{Z}_2^{-\frac{1}{3}}\,mc^2, \tag{36}$$

as already stated.

As E, E_+ increase, δ decreases, and eventually becomes less than δ'. Screening has now an appreciable effect. It may then be shown (B, § 4 (45)) that for complete screening, namely,

$$\frac{\mathrm{EE}_+}{(\mathrm{E}+\mathrm{E}_+)} \gg \frac{137}{2}\,\mathrm{Z}_2^{-\frac{1}{3}}\,mc^2, \tag{45}$$

the differential cross-section is given instead of (32) by

$$d\mathrm{Q} = \frac{8}{\pi}\left(\frac{\mathrm{Z}_1\mathrm{Z}_2}{137}\right)^2\left(\frac{e^2}{mc^2}\right)^2\frac{\mathrm{E}^2+\mathrm{E}_+{}^2+\frac{2}{3}\,\mathrm{EE}_+}{(\mathrm{E}_+ + \mathrm{E})^4}$$

$$\log\,(k137\,\mathrm{Z}_2^{-\frac{1}{3}})\,\log\frac{k'mc^2}{\gamma\,(\mathrm{E}_+ + \mathrm{E})}\,d\mathrm{E}d\mathrm{E}_+. \tag{46}$$

To get the total cross-section roughly, we integrate (34) over all E_T up to $2\,.\,137\,\mathrm{Z}_2^{-\frac{1}{3}}mc^2$, and then integrate (46) from $\mathrm{E}_\mathrm{T} \sim 2\,.\,137\ \mathrm{Z}_2^{-\frac{1}{3}}mc^2$ to mc^2/γ. The result may be written†

$$\frac{28}{27\pi}\left(\frac{\mathrm{Z}_1\mathrm{Z}_2}{137}\right)^2\left(\frac{e^2}{mc^2}\right)^2\log\,(k\,2\,.\,137\,\mathrm{Z}_2^{-\frac{1}{3}})\,\Big\{3\log\frac{k'}{\gamma}$$

$$\log\frac{k''}{2\,.\,137\,\mathrm{Z}_2^{-\frac{1}{3}}\,\gamma} + \log^2(k\,2\,.\,137\,\mathrm{Z}_2^{-}\ )\Big\}, \tag{47}$$

which we may use for $1/\gamma > 2\,.\,137\ \mathrm{Z}_2^{-\frac{1}{3}}$ We have neglected terms like $\log k$ compared with $\log\,(2\,.\,137\,\mathrm{Z}_2^{-\frac{1}{3}})$ consistently with our approximation.

---

† [*Note added in proof, October* 15, 1935. In a recent paper Nordheim (' J. Phys. Rad.,' vol. 6, p. 135 (1935)) has given a formula for the pair creation cross-section with screening, where the term $\log^3 k/\gamma$ of (35) is replaced by $\log^3 (k\ 137\ z_2^{-\frac{1}{3}})$. This is incorrect, as the considerations of § 5 make clear, since only one of the two colliding particles is screened. The correct formula should continue to increase with increasing $1/\gamma$ as $\log^3 k/\gamma$ for very large $1/\gamma$ (as, indeed, (47) does). I have Dr. Nordheim's authority for stating that he agrees with this.]

Since k is a number of the order unity, and $2\,.\,137\,Z_2^{-\frac{1}{3}} > 60$ for all atoms, the error is about 25%. The formula (47) goes over into (35) for $1/\gamma = 2\,.\,137\,Z_2^{\frac{1}{3}}$ to this approximation.

3. In the case when (37) is fulfilled, it may be shown that screening is effective only when

$$\frac{E_+ + E}{c}\,\frac{\gamma^2}{2} \lesssim \frac{Z_2^{\frac{1}{3}}}{137}\,mc, \tag{48}$$

which by (37) becomes for lead, $Z_2 = 82$,

$$\frac{1}{\gamma} \gtrsim 137\,Z_2^{-\frac{1}{3}}\,\frac{(E + E_+)\,\gamma}{2mc^2} > 30. \tag{48A}$$

Screening is therefore appreciable only for very high energies of the particles 1, and its effect is to replace the logarithmic term in (38) by $\log\{k\,137\,Z_2^{-\frac{1}{3}}\,(E_+ + E)\,\gamma/mc^2\}$ which by (48A) is smaller than $\log k/\gamma$.

For a given γ, therefore, such that for the smallest $(E_+ + E)$ subject to (37) the condition (48A) is satisfied, the logarithmic term increases as $\log\{k\,137\,Z_2^{-\frac{1}{3}}\,(E_+ + E)\,\gamma/mc^2\}$ with increasing $(E + E_+)$ until it becomes equal to $\log k/\gamma$, and then remains constant. For larger $(E_+ + E)$, the pair is produced too near the nucleus for screening to have any effect. (*See* §4, 2.) The total cross-section for pairs with $E, E_+ > mc^2/\gamma$ is now

$$Q_3 \sim \frac{1}{3\pi}\left(\frac{Z_1 Z_2}{137}\right)^2\left(\frac{e^2}{mc^2}\right)^2 \log\,(k\,2\,.\,137\,Z_2^{-\frac{1}{3}}), \tag{49}$$

which should be used instead of (39) for $1/\gamma > 137\,Z_2^{-\frac{1}{3}}$.

4—PROBABILITY AS FUNCTION OF IMPACT PARAMETER

The probability of the creation of a pair, the electrons of which have definite momenta, when the particle 1 of charge Z_1 passes at a distance $\mathbf{b}_r$ from the fixed nucleus, is given by (21) with the M's and N's given by (18). The term $\exp\{i\,(\mathbf{p}'_r, \mathbf{b}_r)/\hbar\}$ in the integrals (18) gives the dependence on $\mathbf{b}_r$. For our present discussion it is convenient to choose our axes so that $b_y = 0$. Using (20) and (25), all the integrals (18) are of the form

$$\iint_{-\infty}^{\infty} \frac{1}{\{(\mathbf{p}'_r - \mathbf{p}_{0r} + \mathbf{p}_r)^2 + \delta^2\}\,\{\mathbf{p}'^2_r + \varepsilon^2\}} \times \left[\frac{A}{(\mathbf{p}'_r - \mathbf{p}_{0r})^2 + P^2 - p^2} + \frac{B}{(\mathbf{p}'_r + \mathbf{p}_r)^2 + P_0^2 - p_0^2}\right] e^{ip_x' b/\hbar}\,dp'_x\,dp'_y, \tag{50}$$

where A and B are the corresponding numerators of the quantities **M** or N given by (19).

1. We consider first the case (31), namely,

$$mc^2 \ll E, E_+ \; ; \; mc^2/\gamma \gg E, E_+ .$$

It may now be shown (B, § 2 (16)) that the electron and positron emerge at angles of the order mc^2/E, and mc^2/E_+ respectively, so that by (15)

$$\left.\begin{aligned} \delta &\simeq \frac{(E_+ + E)\, m^2c^3}{2\, EE_+} \\ \varepsilon &\simeq \gamma \frac{E_+ + E}{c} \\ P^2 - p^2 ,\ P_0^2 &- p_0^2 \sim m^2c^2 \end{aligned}\right\} , \quad (51)$$

and hence

$$\delta^2, \varepsilon^2 \ll P^2 - p^2, P_0^2 - p_0^2 .$$

Most of the integral (50) therefore comes from two regions of the order δ and ε round the points $\mathbf{p}_{0r} - \mathbf{p}_r$ and 0 respectively. Suppose $\varepsilon \ll \delta$, then the important region is of the order ε round the point 0, and as b increases from 0 to ∞, the integral will start decreasing rapidly due to the oscillation of the exponential term when

$$b_{\max} \sim \hbar/\varepsilon,$$

i.e., by (51) when

$$b_{\max} \sim \frac{\hbar c}{\gamma (E_+ + E)} . \quad (52)$$

This is for the case $\varepsilon \ll \delta$, *i.e.*,

$$\frac{1}{\gamma} \gg \frac{2\, EE_+}{m^2c^4} . \quad (52\text{A})$$

Thus the creation of pairs of total energy $(E_+ + E)$ takes place up to distances of the order (52) when (52A) holds.

When $\varepsilon \gg \delta$, *i.e.*,

$$\frac{1}{\gamma} \ll \frac{2\, EE_+}{m^2c^4} , \quad (53\text{A})$$

the production of pairs of energy $(E_+ + E)$ takes place up to distance given by

$$b_{\max} \sim \frac{\hbar}{\delta} \sim \frac{\hbar}{mc} \frac{2\, EE_+}{(E_+ + E)\, mc^2} . \quad (53)$$

When $\varepsilon \sim \delta$, $1/\gamma \sim 2\, EE_+/m^2c^4$, and (52) and (53) become the same. We shall see in the next paragraph that these results have a direct physical significance.

2. We now consider the case (37), namely,

$$\frac{mc^2}{\gamma} \ll \mathrm{E}, \mathrm{E}_+ ; \ \frac{\mathrm{M}_1 c^2}{\gamma} \gg \mathrm{E}, \mathrm{E}_+ .$$

It may now be shown that the particles of the pair emerge at angles of the order γ (B, § 3 (28)), so that from (15)

$$\delta \simeq \frac{\mathrm{E}_+ + \mathrm{E}}{c} \cdot \frac{\gamma^2}{2}$$

and

$$\delta^2 \ll \varepsilon^2, \mathrm{P}^2 - p^2, \mathrm{P}_0{}^2 - p_0{}^2 .$$

The important contribution to the integral (50) now comes from a region of order δ round the point $\mathbf{p}_{0r} - \mathbf{p}_r$. The production of a pair, the particles of which have energies E, E_+, therefore takes places at a maximum distance given by

$$b_{\max} \sim \frac{\hbar}{\delta} \sim \frac{\hbar c}{\gamma^2 (\mathrm{E}_+ + \mathrm{E})} . \tag{54}$$

From this we see that if $1/\gamma^2 < (\mathrm{E}_+ + \mathrm{E})/mc^2$, $b_{\max} < \hbar/mc$. In the reverse case it is greater than $\hbar/mc$.

5—CONNEXION WITH PAIR CREATION BY γ-RAYS

We shall now show that there is a very close connexion between the creation of pairs by one charged particle (1) passing through the field of another (2), the process we have been considering above, and the pair creation by γ-rays in the field of the latter particle (2). This is due to the following reason. When the particle 1 (of charge Z_1) moves with a velocity very near that of light, its field suffers a Lorenz contraction, and the electric force perpendicular to its path is $1/\gamma$ times larger than its value when the particle is at rest. To a certain approximation, therefore, the electric field may be considered as perpendicular to the path of the particle. Further, the field has now a magnetic force perpendicular and nearly equal to the electric force, and perpendicular to the direction of motion. In some region, then, small compared with its distance from the path of the particle, the field of the particle is very similar to that of electromagnetic waves, so that we may make a Fourier analysis of the field of the particle and consider each component as a γ-ray. The error in doing this is of the order γ, as has been shown by v. Weizsäcker.† We now use the appropriate formulae, giving the pair production by this γ-radiation in the field of the stationary particle (2) to get the total pair production by the moving particle (1) in the field of the stationary one.

---

† *Cf.* v. Weizsäcker, ' Z. Physik,' vol. 88, p. 612 (1934).

The analysis of the field of the moving particle into monochromatic waves may be schematized according to v. Weizsäcker† thus. If n_ν be the number of quanta of frequency ν per unit frequency range which cross unit area at a distance r from the path of the moving particle 1, then

$$\left.\begin{aligned} n_\nu d\nu &= \frac{2}{\pi}\frac{Z_1^2 e^2}{r^2 c}\frac{d\nu}{h\nu} \qquad && \text{if } \nu < k\frac{c}{2\pi r\gamma} \\ n_\nu d\nu &= 0 && \text{if } \nu > k\frac{c}{2\pi r\gamma} \end{aligned}\right\}, \tag{55}$$

where k is a number of the order unity. To get the total number of quanta N_ν of frequency ν due to a state in which one particle crosses unit area per unit time, we must integrate $n_\nu d\nu$ over all r from $r_{\min}$ to $r_{\max}$. $r_{\max}$ is given by (55), because for larger distances the frequency ν no longer occurs. $r_{\min}$ we take as usual to be the Compton wave-length. We find

$$N_\nu d\nu = 2\pi\int_{r_{\min}}^{r_{\max}}(n_\nu d\nu)\,dr = \frac{Z_1^2}{137}\frac{2}{\pi}\log\left(k\frac{mc^2}{\gamma h\nu}\right)\cdot\frac{d(h\nu)}{h\nu}. \tag{56}$$

To get the effective cross-section for the creation by the colliding particles of a pair, the electron of which has energy E, and the positron an energy E_+, we must multiply $N_\nu d\nu$ by Q (E) dE, the effective cross-section for the creation of a pair, the electron of which has an energy E by a photon of energy $h\nu = (E_+ + E)$ in the field of the stationary particle 2.

We consider first the case (31). Then, according to Bethe and Heitler,‡

$$Q(E)\,dE = 4\frac{Z_2^2}{137}\left(\frac{e}{mc^2}\right)^2\frac{E^2 + E_+^2 + \frac{2}{3}EE_\pm}{(E_+ + E)^3}\left\{\log\frac{2EE_\pm}{mc^2(E_+ + E)} - \tfrac{1}{2}\right\}dE. \tag{57}$$

If we multiply this by (56) we get

$$dQ = \frac{8}{\pi}\cdot\left(\frac{Z_1 Z_2}{137}\right)^2\left(\frac{e^2}{mc^2}\right)^2\frac{E^2 + E_+^2 + \frac{2}{3}EE_+}{(E_+ + E)^4} \times\left\{\log\frac{2EE_+}{mc^2(E_+ + E)} - \tfrac{1}{2}\right\}\log\frac{kmc^2}{\gamma(E_+ + E)}\,dE\,dE_+.$$

This is just formula (32), if we remember that there and here the coefficients of order unity inside the logarithms are indeterminate.

---

† *Cf.* v. Weizsäcker, 'Z. Physik,' vol. 88, p. 612 (1934).

‡ 'Proc. Roy. Soc.,' A, vol. 146, p. 83 (1934).

Formula (57) holds according to Bethe and Heitler if

$$\delta = \frac{(mc^2)^2\,(\mathrm{E}_+ + \mathrm{E})}{2\mathrm{EE}_+}\cdot\frac{1}{c} \gg \frac{mc}{137}\,\mathrm{Z}_2^{\frac{1}{3}}. \tag{58}$$

If E, E_+ are larger, so that the inequality (58) is reversed, screening becomes important, and we must use for Q (E/*d*E)

$$\mathrm{Q}\,(\mathrm{E})\,d\mathrm{E} = 4\,\frac{\mathrm{Z}_2^{\,2}}{137}\left(\frac{e^2}{mc^2}\right)^2 \frac{\left\{(\mathrm{E}^2 + \mathrm{E}_+^{\,2} + \frac{2}{3}\mathrm{EE}_+)\log 183\mathrm{Z}_2^{\frac{1}{3}} - \frac{\mathrm{EE}^+}{9}\right\}}{(\mathrm{E}_+ + \mathrm{E})^3}\,d\mathrm{E}. \tag{59}$$

Multiplying this by (56) gives just (46) if we remember that $\mathrm{EE}_+/9$ in (59) is a term of the order of the error we make in this approximation.

To get the variation with impact parameter *b*, we remark that the maximum distance $r_{\max}$ at which the frequency ν occurs is given by (55)

$$r_{\max} \sim k\,.\,\frac{c}{2\pi\nu\gamma}\,. \tag{60A}$$

Further, according to Bethe and Heitler, a quantum of frequency ν can produce a pair at distances from the fixed nucleus of the order *d* given by

$$d \sim \frac{\hbar}{\delta} = \frac{\hbar}{mc}\,\frac{2\mathrm{EE}_+}{(\mathrm{E}_+ + \mathrm{E})\,mc^2}, \tag{60B}$$

for E, $\mathrm{E}_+ \gg mc^2$. Therefore $b_{\max}$ will be of the order $r_{\max}$ if $r_{\max} \gg d$, *i.e.*,

$$\text{if}\quad \left.\begin{aligned} b_{\max} &\sim k\,\frac{\hbar c}{(\mathrm{E}_+ + \mathrm{E})\,\gamma} \\ \frac{1}{\gamma} &\gg \frac{2\,\mathrm{EE}_+}{m^2c^4} \end{aligned}\right\}, \tag{61}$$

which is identical with (52). Similarly, we get (53) when $r_{\max} \ll d$.

We now consider the creation of slow pairs under the condition (26). According to Nishina and others, one may write $\mathrm{Q}\,(\mathrm{E}_+)\,d\,\mathrm{E}_+$ for E_+, $\mathrm{E} \approx mc^2$ in the form ($\alpha \equiv e^2/\hbar c$)

$$\begin{aligned} \mathrm{Q}\,(\mathrm{E}_+)\,d\mathrm{E}_+ = {} & \frac{1}{64\pi}\,\frac{e^2h}{m^2c^3}\,(\alpha\mathrm{Z}_2)^4\,\frac{1}{(e^{2\pi\alpha\mathrm{Z}_2/\beta_+} - 1)\,(1 - e^{-2\pi\alpha\mathrm{Z}_2/\beta})} \\ & .\Big[2\,(\beta_+^{\,2}\sin^2\theta_+ + \beta^2\sin^2\theta) \\ & + (\alpha\,\mathrm{Z}_2)^2\left(\frac{\pi_2}{8} + 1 + \cos\theta_+\cos\theta\right)\Big]\,\beta_+\,d\beta_+\,d\omega_+\,d\omega, \end{aligned} \tag{62}$$

which is valid for $\beta_+ \equiv v_+/c \ll 1$, $\beta \equiv v/c \ll 1$, $\alpha\mathrm{Z}_2 \ll 1$.

θ_+, θ are the angles of ejection of the positron and electron with respect to the direction of motion of the light quantum.

If, further,

$$\frac{\alpha Z_2}{\beta_+} \ll 1, \frac{\alpha Z_2}{\beta} \ll 1, \tag{63}$$

which is also the condition for the validity of the Born approximation, (62) reduces to

$$\frac{1}{128\pi^3} \frac{e^2 h}{m^2 c^3} (\alpha Z_2)^2 ({\beta^2}_+ \sin^2\theta_+ + \beta^2 \sin^2\theta)\, {\beta_+}^2\, \beta\, d\beta_+\, d\omega_+\, d\omega. \tag{64}$$

This is the formula of Bethe and Heitler calculated by the Born approximation for the case when β_+, $\beta \ll 1$. If we multiply (64) by (56) we get the effective cross-section for the production of slow pairs by the collision of the two particles,

$$\frac{1}{128\pi^3} \cdot \left(\frac{Z_1 Z_2}{137}\right)^2 \left(\frac{e^2}{mc^2}\right)^2 \log\left(\frac{k}{\gamma}\right)^2 \cdot \frac{({p_{+r}}^2 + {p_r}^2)\, {p_+}^2 p^2\, dp\, dp_+\, d\omega\, d\omega_+}{(mc)^8}. \tag{65}$$

This is just formula (27), keeping only the term which predominates as $1/\gamma \rightarrow \infty$, *i.e.*, the logarithmic term. Since (27) is correct for all γ, a comparison of (65) with (27) gives us the error we make by using the Weizsäcker method. For example, the neglect of the term 17/6 in (27) which does not occur in (65) causes (65) to be nearly three times larger than (27), even for $1/\gamma = 10$. For $1/\gamma = 100$, the error is about 20%.

We have thus shown that an adaptation of the Weizsäcker method will to a certain approximation give the same results as the direct calculation. We may then use this method to supplement our results in those regions where the direct calculation cannot be carried out, *i.e.*, for the case $E \sim E_+ \sim 2mc^2$. In this way we may establish that for $E \sim E_+ \sim 2mc^2$, the accurate formula gives a result smaller than that given by (32), which is what was required to justify our method of deriving (34) and (35). In general, one can say that the effective cross-section for the production of a pair, the electron and positron of which have an energy E and E_+ respectively, is got by multiplying the formula (21) of Bethe and Heitler's paper by (56) above. (*Cf.* also their fig. 5.)

In passing, we may remark that a comparison of (62) and (64) shows us the error in the formulae of Bethe and Heitler for slow pairs due to the use of the Born approximation. Integrating (62) and (64) over all angles we have

$$\frac{Q_{\text{Nishina}}}{Q_{\text{Bethe and Heitler}}} = \frac{4\pi^2 (\alpha Z_2)^2/\beta_+\beta}{(e^{2\pi\alpha Z_2/\beta_+} - 1)(1 - e^{-2\pi\alpha Z_2/\beta})} \left[1 + \frac{(\alpha Z_2)^2}{{\beta_+}^2 + \beta^2}\left(\frac{\pi^2}{8} + 1\right)\right].$$

This ratio is less than $\frac{1}{2}$ for

$$\frac{\alpha Z_2}{\beta_+} = \frac{\alpha Z_2}{\beta} = 0\cdot 6. \tag{66}$$

This means that when the condition (63) is not fulfilled, the Born approximation gives a cross-section too large by a factor two. For fast pairs, $\beta_+ \approx \beta \approx 1$, but (63) may still not hold if Z_2 is too large. For lead, instead of (63) we have (66), so that the above analogy seems to indicate that even for fast pairs, where the above argument does not strictly hold, the Born approximation may have given a result too large by a factor two. This would reduce somewhat the discrepancy between theory and experiment found for the stopping of fast electrons. The good agreement between theory and experiment for the production of pairs by γ-rays is, perhaps, to be regarded as fortuitous.

All our results have also been based on the Born approximation, so that our formulae, if applied to lead, may also be too large by about a factor two on this account.

6—RESULTS

The Magnitude of the Cross-Sections—(*a*) We start by giving the cross-section calculated by Heitler and Nordheim† for the creation of pairs by a particle of charge Z_1, mass M_1, moving with velocity small compared with c, on colliding with a particle of charge Z_2 mass M_2, initially at rest. They get

$$Q \approx \left(\frac{1}{137}\right)^2 \left(\frac{e^2}{mc^2}\right)^2 \frac{m^2c^2}{M_1T_1} Z_1{}^4 Z_2{}^2 \left(1 - \frac{M_1Z_2}{M_2Z_1}\right)^2, \tag{67}$$

where T_1 is the kinetic energy of the particle 1. For slow heavy particles, the "first-order process" is far greater than the "second-order," and it is the former which is given by (67). From (67) we see that it *decreases* inversely as the kinetic energy T_1 of the particle creating the pair. If the particle 1 is a proton with $T \sim 10^6$ e.v., and the particle 2 a lead nucleus, $Q \sim 2\cdot 8 \times 10^{-30}$ cm^2. If the particle 1 is an α-particle of the same energy, $Q \sim 1\cdot 3 \times 10^{-30}$ cm^2, the decrease being due to the term $(1 - M_1Z_2/M_2Z_1)^2$. The pair production by slow heavy particles is therefore negligible.

(*b*) We now consider the creation of slow pairs by fast particles subject to the condition (26). The differential cross-section is given by (27). The positron and electron emerge at angles independent of one another, and the probability is greatest for their being ejected perpendicular to the path of the particle 1. The angular distribution is also symmetrical about a plane perpendicular to this path. The integration of this cross-section leads to (34), which increases as the cube of the kinetic energy E_K of the pair. The formula (30) is valid only for $E_K \ll mc^2$. We cannot get an estimate of the partial cross-section for the production of a pair with any kinetic energy less than mc^2 by just integrating (30) from $E_K = 0$ to $E_K = mc^2$, for we make a considerable error in extrapolating (30). We may see this by writing $E_K = mc^2$ in (30) and $E_T = 3mc^2$ in (34), when it appears that (34) is nearly ten times larger than (30). We can, however, use (30) in conjunction with the method of the last section to get a good estimate of the partial cross-section for the creation of a pair with any total energy E_T less than $3\,mc^2$. We find that this partial cross-section is in lead of the order 10^{-28} cm², and 10^{-27} cm², for a proton with $1/\gamma = 2$ and $1/\gamma = 10$ respectively. It increases like (30) logarithmically with the energy of the colliding particle. The cross-section is still extremely small, but is nevertheless about a hundred times larger than the cross-section for pair creation by slow heavy particles.

---

† 'J. Phys. Rad.,' vol. 5, p. 451 (1934).

(*c*) The cross-section for the creation of pairs with energy E, $E_+ \gg mc^2/\gamma$ by heavy particles is given by (38). One sees from this that roughly the probability varies inversely as the fourth power of the total energy of the pair. The mean angles at which the particles of the pair are ejected are of order γ. Also, by integrating (38) over all E_+, we see that the probability of the electron having an energy E, irrespective of that of the positron is proportional to E^{-3}. The total cross-section for the creation of a pair, each particle of which has an energy greater than mc^2/γ, is given by (39), which gives only the order of magnitude. When the particle 1 is a proton and 2 a lead nucleus we get the values given in Table I. The value given in the third row is obtained by using (49) instead of (39), for screening is now effective.

TABLE I—PARTIAL CROSS-SECTION FOR THE CREATION OF PAIRS WITH E, $E_+ > mc^2/\gamma$ BY PROTONS IN LEAD

Energy	$1/\gamma$	Q_3
$\sim 10^9$ e.v.	2	$\sim 2 \times 10^{-27}$ cm²
$\sim 10^{10}$ e.v.	10	$\sim 7 \times 10^{-27}$ cm²
$\sim 10^{11}$ e.v.	100	$\sim 1{\cdot}2 \times 10^{-26}$ cm².

These cross-sections are again about ten times larger than the corresponding cross-sections for the creation of slow pairs by protons of the same energy, so that if a proton produces a pair at all it will be a fast pair. The effect is still very small.

(*d*) The cross-section for the creation of pairs subject to the condition (31) is given by (32) and (46). The distribution of energy between the particles of the pair, and the dependence of the probability on the total energy of the pair is almost exactly the same as for the creation of the same pair by a γ-ray, multiplied by the factor $\log\{mc^2/\gamma(E_+ + E)\}/(E_+ + E)$, as indeed the considerations of § 5 make clear. The term $\log\{mc^2/\gamma(E_+ + E)\}$ varies but little. For pairs with $E_T \equiv (E_+ + E)$ large, *i.e.*, greater than $100\, mc^2$, there is a tendency for one particle to get more energy than the other. This tendency increases as the total energy E_T increases. For $E_T \sim 20\, mc^2$ there is a very broad maximum when both the particles get about the same amount of energy, and this maximum becomes more pronounced as E_T decreases. The distribution curves are like those given in the paper by Bethe and Heitler (p. 107). The mean angles at which the electron and positron appear are roughly mc^2/E and mc^2/E_+ respectively, and the two particles tend to emerge on opposite sides of the z axis. The effect of screening is also exactly the same as for γ-rays, and has been considered in § 3.

The integration of (32) over all energies of the positron E_+ gives, under the condition $mc^2 \ll E \ll mc^2/\gamma$, just the expression (34) with E substituted for E_T. This shows that the probability of the electron having an energy E, independently of the energy of the positron, is roughly proportional to $1/E$.

The total cross-section is roughly given by (35) for $1/\gamma < 2\,.\,137\, Z_2^{-1/3}$, and by (47) for $1/\gamma > 2\,.\,137\, Z_2^{-1/3}$. The resulting values of Q for different energies are given in Table II. Another quantity of interest is Q/Q_γ where Q is the cross-section for the creation of a pair by an electron, and Q_γ is the cross-section for the creation of a pair by a γ-ray of the same energy. This ratio is approximately

$$\frac{Q}{Q_\gamma} \sim \frac{1}{3\pi}\cdot\frac{1}{137}\,Y,$$

where

$$Y = \begin{cases} \log^2 1/\gamma & \text{for } 1/\gamma < 2\,.\,137\, Z_2^{-1/3} \\ \left\{3 \log\dfrac{1}{\gamma} \log\dfrac{1}{2\,.\,137\, Z_2^{-1/3}\,\gamma} + \log^2 2\,.\,137\, Z_2^{-1/3}\right\} & \text{for } 1/\gamma > 2\,.\,137\, Z_2^{-1/3}. \end{cases}$$

It is given in the third row of Table II.

TABLE II—CROSS-SECTIONS IN CM² FOR THE CREATION OF PAIRS IN LEAD

$1/\gamma$	10	50	100	500	1000
Q	11×10^{-26}	56×10^{-26}	$0\cdot9 \times 10^{-24}$	$2\cdot2 \times 10^{-24}$	$2\cdot9 \times 10^{-24}$
$Q/Q_\gamma \sim$	(0·004)	0·01	0·02	0·05	0·06

The cross-sections up to $1/\gamma = 100$ apply to protons as well as electrons, and will be relatively accurate for them. $1/\gamma = 100$ corresponds to a proton of $\sim 10^{11}$ e.v. For electrons of energy $10mc^2$ the value of Q given is likely to be too large, since for such low energies the deflection of the electron by the nucleus as well as the reaction of the electrons of the created pair may no longer be neglected. The values become progressively more accurate with increasing energy.

The Effect of Exchange—We shall now consider qualitatively the effect of exchange, when the particle creating the pair is an electron. Our calculations are accurate only in the region $E, E_+ \ll mc^2/\gamma \equiv E_i$ and $E_i \gg mc^2$.

E_i denotes the initial energy of the electron 1. The electron after the pair creation is left with an energy E_f very nearly given by $E_f \simeq E_i - E_T$, since in all important cases the energy communicated to the particle 2 is small compared with mc^2. Since $E_T \ll E_i$ by our assumptions, $E_f \sim E_i \gg E$. Let $P(p_i \rightarrow p_f, p_0 \rightarrow p)$ be the matrix element without exchange for the transition in which the electron 1 jumps from a state of momentum p_i to a state of momentum p_f, and the electron of the created pair jumps from its initial state p_0 to a final state p. Then $|P(p_i \rightarrow \rho_f, p_0 \rightarrow p)|^2$ integrated over all final states p_f of the electron 1 is just the process we have calculated, and is given by (32) under the conditions (31). Let $P(p_i \rightarrow p, p_0 \rightarrow p_f)$ denote the matrix element for the process in which the electron 1 jumps from its initial state p_i to a final state p, and the electron of the created pair jumps from an initial state p_0 to a final state p_f. Then $|P(p_i \rightarrow p, p_0 \rightarrow p_f)|^2$ integrated over all final states p is given very roughly by (32) if we there write E_f instead of E, since (32) is not accurate when the electron creating the pair loses a large fraction of its initial energy. More accurately, $|P(p_i \rightarrow p, p_0 \rightarrow p_f)|^2$ would be still smaller. Hence $|P(p_i \rightarrow p, p_0 \rightarrow p_f)| \ll |P(p_i \rightarrow p_f, p_0 \rightarrow p)|$ if $E_i \sim E_f \gg E$. If we had taken exchange into account accurately, then the cross-section for the process in which one electron was in a final state p_f, and the other in a final state p would have been proportional to

$$|P(p_i \rightarrow p_f, p_0 \rightarrow p) - P(p_i \rightarrow p, p_0 \rightarrow p_f)|^2,$$

which, on account of the inequality given above, is very nearly $|P(p_i \rightarrow p_f, p_0 \rightarrow p)|^2$. For those cases where $E_f \sim E$, both matrix elements are of the same order, and the effects of exchange are not so simple. We may thus sum up by saying that exchange has but a small effect on formulae (27), (30), (32), and (34) in the region where they are valid, *i.e.*, $E_T \ll mc^2/\gamma$.

Cross-Section as Function of the Total Energy of the Pair—For a given large energy of the particle creating the pair, the differential cross-section as function of the energies of the particles of the pair behaves as follows. For $p, p_+ \ll mc$, the cross-section for the creation of a pair of total kinetic energy E_k increases as $\mathrm{E}_k{}^3$, as shown by (30), reaches a maximum for some total energy E_T of the order $3mc^2$, and then for $\mathrm{E}_\mathrm{T} \gg mc^2$ decreases nearly as $\mathrm{E}_\mathrm{T}{}^{-1}$, as seen from (34). If the particle creating the pair be a heavy particle, then the further case $\mathrm{E}_\mathrm{T} \gg mc^2/\gamma$ is possible, and in this region the cross-section decreases approximately as $\mathrm{E}_\mathrm{T}{}^{-3}$, as one may see by integrating (38).

Comparison of Pair Creation by Protons and Electrons—One sees from the above cross-sections that if a particle of energy 10^{10} e.v. is observed to create a pair in its passage through a plate, the chance of its having been a heavy particle, *i.e.*, with rest mass comparable to that of the proton, is less than 4%, whereas for a particle of 10^9 e.v. the chance is less than 0·2%. Further, the angles at which the particles of the pair emerge are much larger for creation by protons than by electrons. For example, for a particle of $\sim 10^9$ e.v. γ, if it be a proton, is 1/2. For a pair with $\mathrm{E}, \mathrm{E}_+ \gg mc^2/\gamma \sim 2mc^2$, the mean angles at which the electrons of the pair emerge are roughly γ. If the particle creating the pair had been an electron, the mean angles would have been of the order $\dfrac{mc^2}{\mathrm{E}}, \dfrac{mc^2}{\mathrm{E}_+}$ respectively, which are much smaller than γ.

In conclusion, we remark that the errors which may be caused by the use of the Born approximation are exactly the same here as in the creation of pairs by γ-rays, calculated by Bethe and Heitler. For the reasons given in § 5, the results for lead for slow pairs are liable to be too large roughly by a factor two on this account.

SUMMARY

The creation of electron pairs in the collision of particles moving with relative velocity very near the velocity of light is calculated. The effect of screening is considered, and the variation of probability of pair creation as function of impact parameter is investigated. It is shown that to a certain approximation most of the formulae can be derived by a method similar to one due to v. Weizsäcker, where the field of the moving particle is considered as a superposition of γ-rays.

The effective cross-section for the pair creation by fast protons in lead is

more than a thousand times larger than the cross-section for pair creation by slow protons calculated by Heitler and Nordheim. For the collision of electrons of 10^8 e.v., with a lead nucleus, the cross-section is of the order 10^{-24}cm^2, and is about one-fifteenth of the cross-section for the creation of a pair by a γ-ray of the same energy. It increases with increasing energy of the electron.

On the Instability of a Fluid when Heated from Below

By R. J. SCHMIDT, Ph.D., and S. W. MILVERTON, Ph.D., D.I.C.

(*Communicated by S. Chapman, F.R.S.—Received June* 7, 1935)

[PLATE 15]

INTRODUCTION

It has been known for some time that when a horizontal layer of fluid is heated from below the fluid remains stationary, if initially at rest, until a certain temperature difference, depending on the physical constants of the fluid and the depth of the fluid layer, is reached. The equilibrium, which becomes less and less stable as the temperature rises, then becomes unstable for infinitely small disturbances, so that the fluid begins to move.

Rayleigh,[*] in 1916, put forward a theory which gave the temperature at which motion first occurs when the top and bottom layers are free surfaces. The results agreed qualitatively with previous experiments due to Bénard.[†] Later Jeffreys[‡] calculated a critical temperature for the stability of a fluid between two rigid horizontal conducting planes. The investigation has since been revised by Jeffreys[§] and also by Low.[¶]

The mathematical work involves certain not yet fully established assumptions, and it appeared desirable to investigate the problem experimentally to test the theoretical result. The mathematical investigations have not so far shown whether any sudden increase occurs in the heat transfer when motion begins, but this appeared probable and is also investigated in the present work. It was decided to suspend two plates

[*] 'Phil. Mag.,' vol. 32, p. 529 (1916).

[†] 'Ann. Chim. Phys.,' vol. 23, p. 62 (1901).

[‡] 'Phil. Mag.,' vol. 2, p. 833 (1926).

[§] 'Proc. Roy. Soc.,' A, vol. 118, p. 195 (1928).

[¶] 'Proc. Roy. Soc.,' A, vol. 125, p. 180 (1929).

horizontally in water, to heat the lower plate electrically and to measure the temperatures of the plates by means of thermocouples.

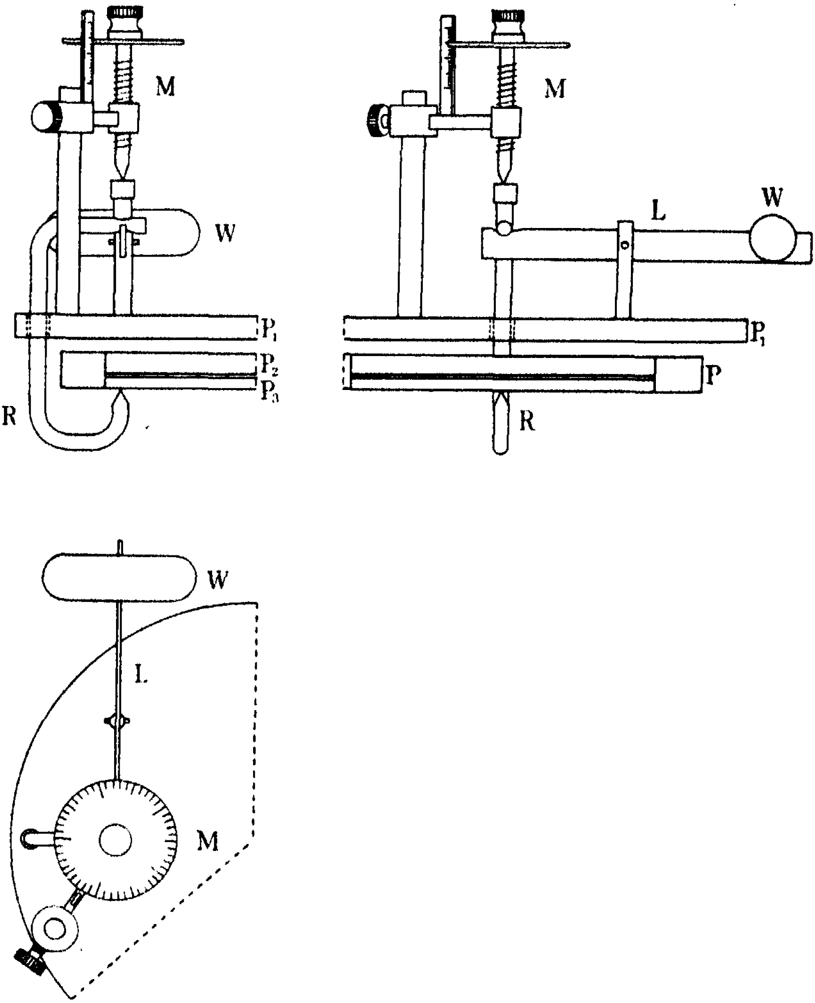


FIG. 1.

DESIGN OF APPARATUS

The apparatus employed, of which a diagram is given in fig. 1, was designed so that two circular brass plates, accurately parallel and separated by a known distance, could be maintained at a known temperature difference when placed in water. Care was taken to eliminate edge effects and to ensure temperature equality over their surfaces.

The plates were cast in brass, the thickness of each being $\frac{3}{8}$ inch. The diameters of the upper and lower plates P_1 and P_2 were 8 and 5 inches respectively. The surfaces facing each other were rendered as nearly plane as possible, a surface plate being used for comparison. To the lower side of the small plate was screwed a similar plate P_3, a flat heating coil being sandwiched between the two. This coil was of eureka wire, wound uniformly on mica, and had a resistance of about one ohm. A $\frac{3}{4}$ inch ring of paraffin wax P was built up round the edge of the composite plate. Eureka-copper thermocouples were inserted in holes drilled from the edges of both plates to near their centres.

It was desirable that no disturbance—thermal or mechanical—should occur in any part of the liquid as this might cause instability in the liquid between the plates at a temperature below that given by the theory which assumed that the disturbances were infinitesimal. For this reason, it was necessary that the rigid connexions between the upper and lower plates should be made of a thermal insulator. The lower movable plate was therefore supported at three points by glass rods R, of which the movement relative to the upper plate could be measured. Each glass rod R was bent at right angles three times, and passed loosely through a hole in the top plate. The weight was supported by one arm of a lever L, on the other arm of which was a weight W which could be moved along the lever. To prevent strain on the glass where it passed through the plate, it was arranged that the lever fitted into a slight depression in the rod, vertically over the point of support, the rod being prevented from turning by means of a V-shaped cut in the lever. Above the point of support a short piece of glass rod was fused vertically to the main rod, and fitted with a brass cap. The position of the lower brass plate was varied and determined by means of a micrometer screw M, rigidly supported on a thick brass rod screwed into the upper plate, the point of the screw making contact with the cap. By this arrangement, the plate was supported independently of the micrometer, and by adjustment of the position of W, the strain on the screw could be made as small as desired. The top plate rested on three vertical 4-inch glass rods fixed in a wooden base.

EXPERIMENTAL PROCEDURE

The apparatus stood in a tank which was supported on three levelling screws. Water was introduced until almost level with the upper surface of the top plate, care being taken to eliminate all air bubbles between the plates. The micrometer readings were taken with the plates in contact, and adjusted so as to give the required separation.

After allowing a suitable time for constant conditions to obtain, a small heating current was passed, and the temperatures of the plates taken when the difference had become constant. The current was increased by steps every ten minutes and the corresponding pairs of temperatures were found. Results were obtained for each of four separate distances of separation.

TABLE I

Experiment 1
$d = 4$ mm

C	θ_1	θ_2
1·48	22·10	22·60
1·85	22·40	23·12
2·22	22·70	23·70
2·59	22·87	24·22
2·96	23·20	24·90
3·35	23·60	25·58
3·70	24·05	26·22
4·07	24·27	26·80
4·81	25·20	28·45
5·93	26·32	30·70

Experiment 2
$d = 4·5$ mm

C	θ_1	θ_2
0·56	22·15	22·27
0·95	22·20	22·45
1·30	22·35	22·75
1·57	22·47	23·05
1·70	22·57	23·25
1·87	22·65	23·42
2·07	22·75	23·72
2·33	22·87	24·00
2·59	22·97	24·25
2·89	23·22	24·80
3·24	23·50	25·32
3·69	23·75	25·85
4·17	24·17	26·67
4·63	24·85	27·72
5·07	25·25	28·50
5·56	26·00	29·82

Experiment 3
$d = 5$ mm

C	θ_1	θ_2
0·74	19·57	19·75
0·93	19·55	19·85
1·20	19·55	19·95
1·48	19·57	20·15
1·76	19·62	20·37
2·04	19·75	20·72
2·31	19·87	21·05
2·63	20·02	21·40
2·96	20·27	21·85
3·30	20·50	22·35
3·61	20·77	22·85
3·83	21·05	23·30

Experiment 4
$d = 5·5$ mm

C	θ_1	θ_2
0·56	17·00	17·07
0·74	17·00	17·15
0·93	17·10	17·32
1·11	17·17	17·47
1·30	17·27	17·67
1·48	17·35	17·90
1·67	17·47	18·15
1·85	17·62	18·45
2·06	17·75	18·72
2·22	17·95	19·07
2·41	18·12	19·35
2·59	18·35	19·70
2·78	18·55	20·02
2·98	18·72	20·35
3·15	18·92	20·62

RESULTS

The results of four experiments are given in Table I, in which C is the heating current in amperes, θ_1 and θ_2 are the respective temperatures of the upper and lower plates in degrees C, and *d* is the distance between the plates.

DISCUSSION OF RESULTS

Owing to the indeterminate nature of the conditions below the heated plate, we are unable to predict exactly how the heat loss from the plate depends on the temperature measured, even if pure conduction is assumed. It is, however, probable that if there is a sudden change in the method of heat transmission in the layer of liquid between the plates, a discontinuity will be evident in the curve connecting the energy supplied and the temperature difference between the plates.

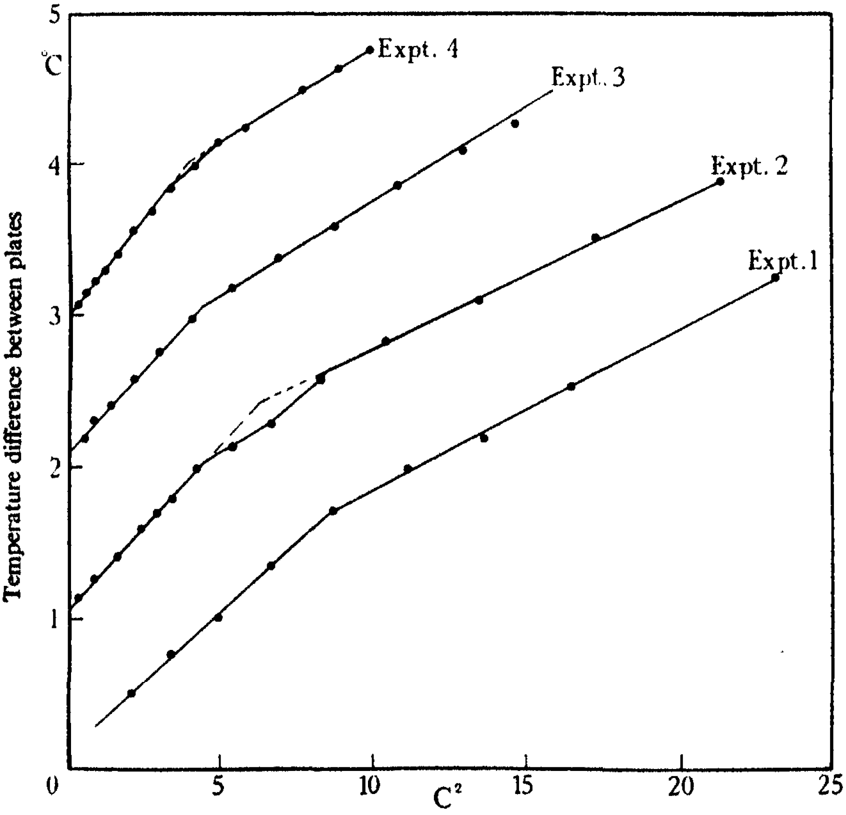


FIG. 2—Curve showing the discontinuity in the rate of heat transfer in a layer of fluid between two horizontal plates when heated from below. The origins of Curves 2, 3, and 4 are displaced by 1, 2, and 3 degrees respectively up the temperature difference axis.

Fig. 2 shows the relation between the heat input (which is proportional to the current squared) and the temperature difference between the plates, for each of the experiments of which the readings are given in Table I. It can be seen from the graphs that the relation is linear for pure conduction and that there is then a change-over to another linear law. It was assumed that instability first occurred at the point where the change-over began.

If a fluid under uniform vertical temperature gradient is at rest between two horizontal perfectly conducting planes, then, according to Jeffreys's revised theoretical result (*loc. cit.*), instability will occur when $\lambda = 1709$ where λ is defined by

$$\lambda \equiv \frac{g\,\alpha\,c\,\rho\,(\theta_1 - \theta_2)\,d^3}{k\,.\,\nu},$$

where g is the acceleration due to gravity, and k, ν, c and α are respectively the thermal conductivity, the kinematic viscosity, the specific heat, and the coefficient of expansion of the fluid. The analysis assumes that k, ν, c, and α are independent of the temperature. This is not true of real fluids, and it is a matter of considerable difficulty to extend the analysis so as to cover the variation of these quantities with temperature. In consequence of this variation, λ varied by as much as 12% through the water in some of the experiments.* For each point of instability, the value of λ at the mean temperature of the water was calculated and is given in Table II.

TABLE II

	Experiment			
	1	2	3	4
Temperature of upper plate	23·2	22·75	19·80	17·63
Temperature of lower plate	24·9	23·75	20·85	18·47
Value of λ at mean temperature	1970	1580	1850	1670

The mean result is 1770 and the average departure from this mean in the individual experiments is 140. Thus, within the limits of experimental error, the values agree with the theoretical value, 1709.

OPTICAL METHOD

An optical method of investigating the heat loss from a surface has recently been described by Saunders, Fishenden, and Mansion,† who kindly lent their apparatus so that the method might be used in the present work.

---

* There is a certain amount of uncertainty as to the value of the thermal conductivity of water. The value used was 0·00145 cal/cm/sec/°C, this being the mean of the values obtained by Lees, by Weber, and by Milner and Chattock. (Kaye and Laby, "Physical and Chemical Constants," p. 55 (1928).)

† "Engineering," vol. 139, p. 483 (1935).

The apparatus stood in a large tank having optical glass sides. Light from a distant arc was passed between the plates, grazing the lower one. The light, limited by an adjustable horizontal slit, was received on a vertical screen several yards distant from the tank. The usual procedure for a set of readings was followed, the image on the screen being watched.

It was found that a change in the character of the image occurred at the same time as the change in the slope of the input-temperature curve. Thus the optical method gave an additional verification of the theory giving the conditions for instability; moreover, interesting data were obtained regarding the motion of the fluid after the instability point had been reached.

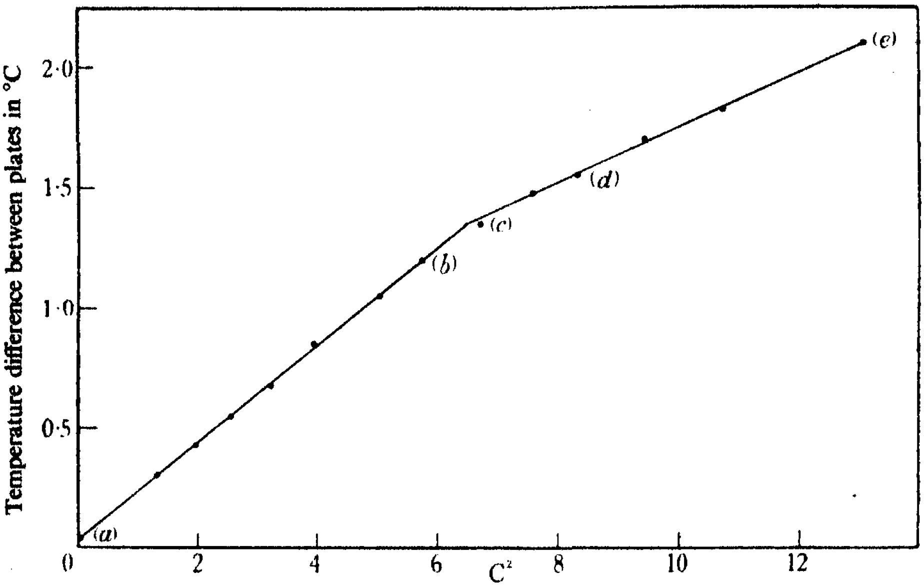
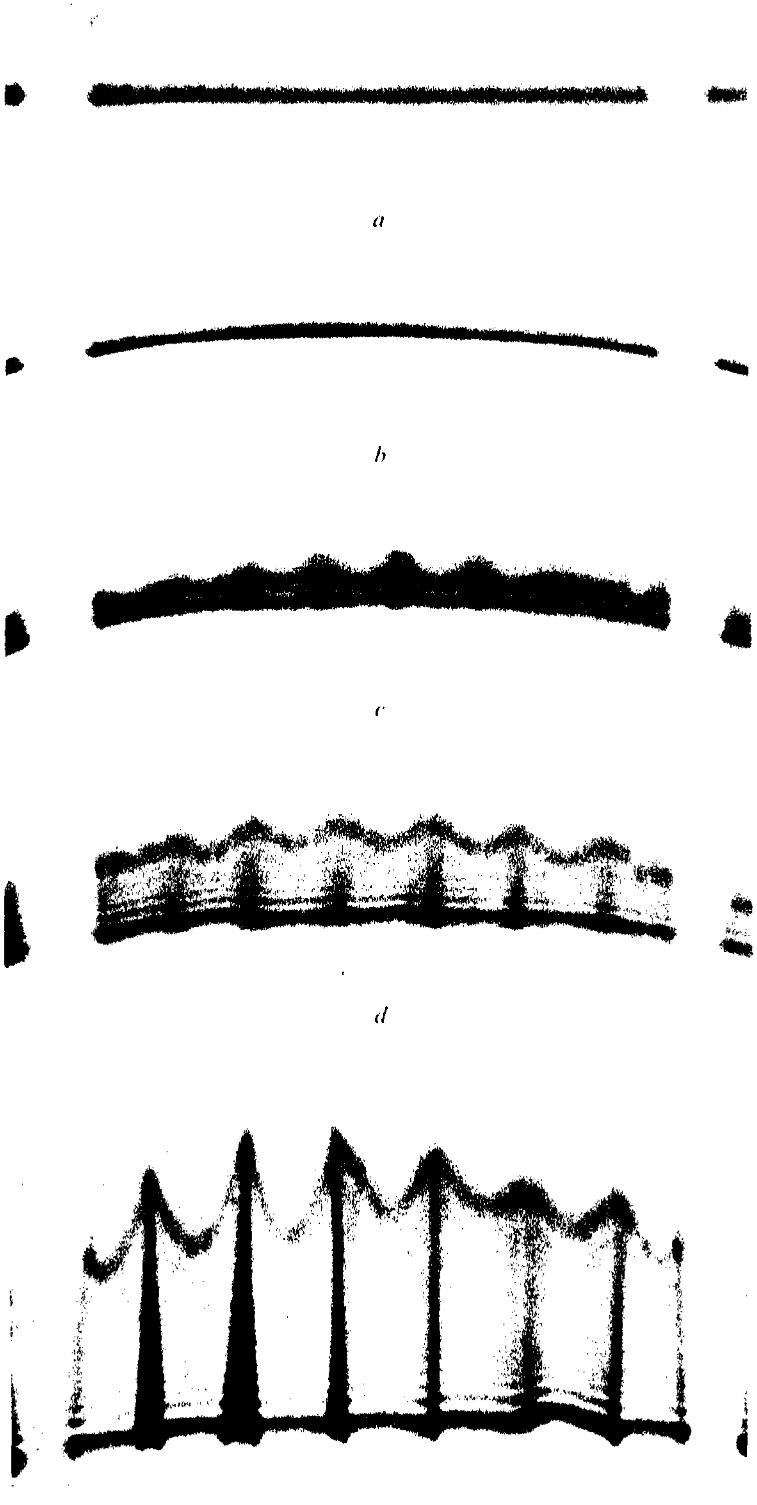


FIG. 3—$d = 5$ mm.

The results of an experiment are plotted in fig. 3, while the negative photographs (*a*) to (*e*) in fig. 4, Plate 15, were taken at the screen at times corresponding to the indicated points in fig. 3.

(*a*) This photograph shows the undeflected image of the slit, no heating current being passed. The breaks in the band are due to the glass supports of the upper and lower plates.

(*b*) A temperature gradient has been established but the instability point has not been reached. Under ideal conditions the image would be a semi-ellipse, since, the temperature gradient being everywhere the same, the vertical deflection of a ray would be proportional to the chord of the circular plate along which it travelled. Except for a slight edge effect, due,



a

b

c

d

e

FIG. 4.

no doubt, to the upper plate being larger than the lower, the image is seen to have approximately this shape.

(*c*) The instability point has now been passed, and it is seen that the band has widened and its upper edge has become wavy, while there are indications of bright vertical lines corresponding to the crests. These effects can easily be explained if we assume that convection is taking place and that the motion of the liquid is cellular, as is predicted by the theory.

The curvature of a light ray depends on the temperature gradient perpendicular to the direction of the ray, and, for steady temperature conditions, the vertical gradient very near the surface of the lower plate, where conduction alone is operative, is greater than in the moving liquid above. Hence the deflection of a ray which grazes the surface does not depend—for a given heat transfer—on whether the liquid above is stationary or in motion, whilst a ray which passes through the moving liquid is deflected less. Hence the image on the screen is widened by the upper edge crossing the lower and approaching the zero position. This was seen to occur fairly quickly when the instability point had been passed, and is evident in the photograph.

If the liquid is in convective cellular motion, it is clear that there will be horizontal temperature gradients in it. Thus rays will be deflected horizontally towards the cooler descending parts of the liquid, and the cells, if arranged regularly, will produce a series of vertical brighter patches, as in the photograph.

At points where the cooler liquid is descending, the lower plate is losing heat slightly faster than where the warmed liquid is rising, and the consequent slight variation in the vertical temperature gradient at the lower plate causes the upper edge of the image to be wavy, as is seen. The crests correspond to the descending parts of the liquid and therefore with the bright vertical patches.

It is noteworthy that the image was found to be quite steady, and could be kept the same almost indefinitely.

The distance between the vertical bright patches was 1·8 cm, and it might be thought that the size of the cells could be calculated from this result.* But there are several difficulties in the way. The light is bent in a complicated manner when passing through the fluid. We do not know the angle at which the light strikes the edges of the cells, nor do we know the horizontal cross-section of the cells. This cross-section is probably square near the centres of the plates but may not be so in other parts of the fluid.

---

* It is worth recording that when these photographs were taken, the distance of the screen from the arc was 76 feet and that the distance of the apparatus from the arc was 43·5 feet.

(*d*) This photograph is similar to the preceding one, but the upper edge of the image is deflected farther from the zero position on account of the greater heat transfer. The image is wider than in (*c*), the movement of the liquid being evidently more violent, while the distance between successive vertical lines is 2·1 cm.

(*e*) The heat transfer is now greater still, and the motion correspondingly violent. The distance between crests is now 2·2 cm. The development of some of the vertical lines into V-shaped forms is an effect which was usually noticed when the temperature gradient was much greater than that required for instability. The explanation is not at first sight apparent, but may be concerned with the screen not being in the best position for horizontal focussing of the light passing through different parts of the cells.

The authors wish to thank Professor C. H. Lander, of the Mechanical Engineering Department, Imperial College, in whose laboratory the work was carried out, and Mr. O. A. Saunders and Mr. H. D. Mansion for help received during the optical part of the experiment.

SUMMARY

Two brass plates were supported parallel to one another in a glass tank containing water and the lower plate was heated electrically. The water remained at rest until a certain temperature difference was reached, and then convection began between the plates. The temperature difference at which motion began agreed, within the limits of experimental error, with a formula predicted from theoretical considerations by Jeffreys. Graphs of the temperature difference between the plates plotted against the rate of heat transfer show a sudden break in direction when instability occurred.

Light from a point source was passed through the water and the change in appearance of the image on a screen also showed when instability occurred.

Progressive Lightning—II

By B. F. J. SCHONLAND, M.A.,* Ph.D., D. J. MALAN,* M.Sc., D.ès.Sc., and H. COLLENS,† M.(S.A.)I.E.E.

(*Communicated by Sir Charles Boys, F.R.S. Received June* 11, 1935)

[PLATES 16-21]

1—INTRODUCTION

An account has previously been given of some results obtained with the Boys camera in the study of lightning discharges and of some deductions as to the nature of these discharges.‡ The present paper is concerned with further studies in the same field with improved technique and apparatus.

The material now available includes 41 photographs comprising 95 separate flashes and over 200 lightning strokes. These have been obtained from 13 different thunderstorms over a period of three years, and may therefore be considered to offer a fair sample of the lightning discharges of Southern Africa. We have derived from them a general picture of the nature of the discharge and the manner of its development which seems to be valid for the great majority of the cases studied.

This would suggest that the picture is applicable to lightning observations elsewhere, in temperate as well as tropical regions. It is therefore of interest to note that the results described in the previous paper have been confirmed by observations made in the Northern United States (lat. 42° N.) by K. B. McEachron and others§ with a Boys camera. Further support for this view is to be obtained from the early observations of Walter,‖ whose work with a slow-moving single-lens camera is well known.

It is thus probable that the interpretation of our results which follows applies to flashes from storms in temperate as well as tropical climates.

---

* The University of Cape Town.

† The Victoria Falls and Transvaal Power Co., Johannesburg.

‡ Schonland and Collens. 'Proc. Roy. Soc.,' A, vol. 143, p. 654 (1934), referred to as I.

§ 'Elect. (Cl.) J.' vol. 31, p. 251 (1934). (I am much indebted to Mr. McEachron and other members of the staff of the General Electric Co. for the opportunity of examining some of these photographs.—B. S.)

‖ 'Ann. Physik,' vol. 10, p. 393 (1903), and 'Jahrb. Hamburg wiss. Anstalten,' vol. 20 (1903). We regret that these observations and deductions escaped our notice until recently. They are further discussed in §§ 6 and 7.

In temperate regions, however, a type of flash which is rare in South Africa and perhaps not so rare elsewhere—discharge from a positively charged cloud volume—may exhibit interesting and important differences in development.

The present paper is devoted to an analysis of the main facts which are likely to enter into any general discussion of the lightning discharge and its effects. The theoretical questions raised will shortly be dealt with in a further publication, as also will a few exceptional discharges presenting unusual features.

It is necessary to point out that practically all the discharges which have been photographed occurred between the base of a thundercloud and the earth. Some information has been obtained about discharges which fail to reach the ground (§ 10), but the most frequent type of discharge, that taking place entirely within the cloud, cannot be dealt with by these methods at all.

2—APPARATUS AND TECHNIQUE

For successful photography with the Boys camera, it is necessary to operate the device in the country, away from the lights of towns, for the camera must be exposed for some time before a photograph is obtained. It is therefore preferable to drive the rotating lenses by hand, and the difficulty of determining the speed of rotation is solved by operating just beyond one of the critical frequencies of vibration of the camera.

The type of camera used in most of the present work has already been described (I, p. 654). The speed of rotation was about 3000 revs/min, nearly twice that previously employed. With good conditions this means a resolving power in time of 0·6 microseconds (I, p. 656). For flashes 12 to 95, the photographs from the Boys and the fixed camera were supplemented by others from a slow-moving camera, working off the same drive as the fast Boys instrument. This camera consisted of a single lens, making one revolution for every 59 of the two-lens instrument. It was originally provided to give the order of appearance of the successive strokes of the discharge, but turned out extremely valuable in that it clearly showed the slower leaders to strokes.

A form of Boys camera in which the two lenses are fixed and the film revolves on the inside of a drum has recently been employed by us (flashes 82 to 87).* This arrangement gives resolving powers as low as 0·3 microseconds, since the relative velocity of the two images of the discharge is 35 metres/sec. This form of Boys camera has several advantages, for it gives a linear record instead of a circular one, and the length of the linear track is 2·3 times as great. There is consequently less overlapping of the stroke images and less fogging from cloud discharges.

---

* Constructed by Ross and Co. to the design of Sir Charles Boys, and slightly modified. This camera was very kindly presented to the South African Institute of Electrical Engineers by Mr. Bernard Price, O.B.E. For the fitting of the driving gear and auxiliary cameras we wish to thank Mr. C. J. Monk and Mr. F. C. A. Crewe.

On many of the later photographs, *e.g.*, flash 76, fig. 13, Plate 20, one lens has been open to *f*/6·3 and the other stopped down to *f*/16. This arrangement allows of photometric study of the relative intensity of the light from different parts of a stroke, the leader and the main return portion, for example.

In discussing the records we have, as before, allotted to each flash a number and distinguished the strokes forming one flash by letters. After flash 11, these letters correctly indicate the actual place of the stroke in the series; before flash 11, since no slow camera was available, they were allotted at random.

3—GENERAL ACCOUNT OF THE NATURE OF FLASHES TO GROUND

In this section an account is given of the general features of the lightning discharge which are common to most of the photographs studied, the evidence for the picture presented being discussed later.

Each lightning flash to ground consisted of a series of separate strokes ranging in number from 1 to 27. The time of separation was very variable, even in the same series, and values of from 0·0006 sec to 0·53 sec have been found.

The first stroke of each series differs from the others in two respects, which make it easy to identify. It is usually much more intense than the succeeding strokes and is always much more heavily branched.

Each stroke is composite, and consists of a leader portion travelling from cloud to ground and a main return portion moving faster and more brightly in the reverse direction. The leader to the first stroke consists of a series of streamers moving downwards in a step-by-step manner, and is hence referred to as "*stepped.*" The length of each step is about 50 metres, and after completing a step the tip of the streamer appears to pause for a time of the order of 100 μ-sec,* whereupon the streamer extends still further, the new step being much brighter than the rest of the streamer.

The prolific branching of the main part of the first stroke of a series arises solely from downward branching in the stepped leader which precedes it. The course followed by the steps decides the path which is followed generally by all the strokes making up the flash and the zig-zag nature of the channel arises from changes in the direction adopted by successive steps.

---

* μ-sec = microseconds.

The leaders to the *subsequent strokes* of the series are in general of the dart-like character previously described. They follow the path blazed by the first leader but usually do not branch. When branches are formed by either first or successive leaders they are subsequently more brilliantly illuminated during the return stroke.

Occasionally, however, when there has been an unusually long interval before the occurrence of a "subsequent" stroke, the lower end of the usual dart leader becomes stepped like the leader to a first stroke. This type of leader is termed *dart-stepped*.

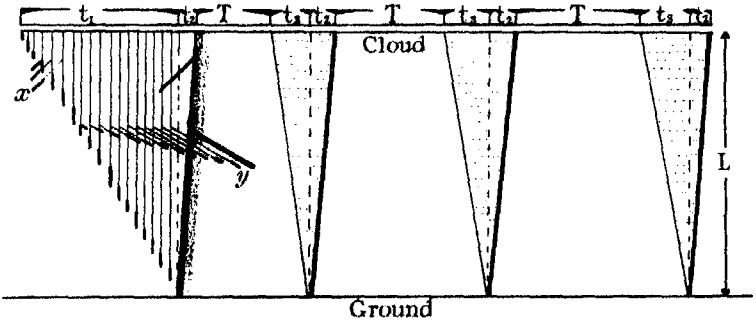


Fig. 1—Average values: t_1 . . 0·01 sec; t_2 . . 0·001 sec; t_3 . . 0·00004 sec; T . . 0·03 sec; L . . 2·0 km.

The second, upward-moving, return portion of each stroke follows directly upon the arrival of the leader part at the ground. In the course of its journey the return part branches outwards and downwards along the forks blazed by the leader. In some flashes, as at y in fig. 1, the leader is still moving out along a branch when the return part catches it up. In others, as at x in fig. 1, the leader branch has ceased to develop some time before the return portion starts. The very bright luminosity of the return stroke has its longest duration and greatest intensity at the ground and often shows a marked decrease in intensity at successive branches from the bottom upwards.

The main features of a straight vertical discharge as they would be shown on an endless film by one lens of a Boys camera with horizontal displacement are represented diagrammatically in fig. 1, which is not drawn to scale. The mean values of the various time intervals make it clear why several camera speeds are necessary to obtain the fullest information about the discharge. The Boys method of studying the progress of a stroke depends, of course, upon the use of two records like fig. 1 with lens displacements in opposite directions.

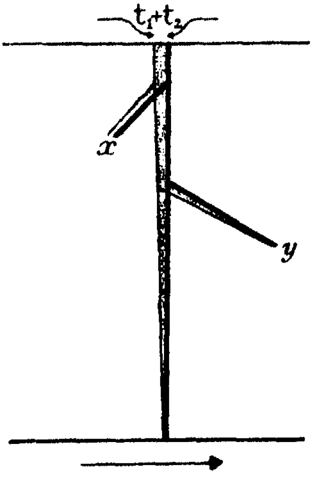


FIG. 2.

The type of record given by the slow-moving camera for the first stroke of fig. 1 is illustrated in fig. 2. Results of this kind were first obtained by Walter (*loc. cit.*) with a slow single-lens camera in 1902, and similar effects are to be found on photographs taken by Larsen* in the same way. In both cases, however, the camera speeds were too slow to give the stepped-leader processes of fig. 1.

A good example of the usefulness of the slow camera in giving the order of strokes is afforded by the two records of flash 75, which consisted of seven separate strokes, shown in fig. 3, Plate 16.

4—GENERAL DATA ON THE DISTRIBUTION OF STROKES IN NUMBER, TIME, INTENSITY, AND MANNER OF BRANCHING

(*a*) *Number*—The total number of separate strokes making up the 65 flashes to ground which are available for this investigation was 235, an average of 4 per discharge. The distribution is shown in Table I, which indicates that single-stroke flashes were the most frequent.

TABLE I

No. of strokes per flash....	1	2	3	4	5	6	7	> 7 (8, 8, 9, 10, 13, 15, 27)
No. of cases recorded	32	5	4	5	8	2	1	8

I am indebted to Dr. Abbott, Secretary of the Smithsonian Institution, for the opportunity to examine these photographs.—B. S.

(*b*)—*Time Intervals between Strokes*—The time intervals between the separate strokes of a series varied from 0·001 to 0·53 seconds. The most frequent value for this interval was 0·03 seconds, as shown by the distribution Table II.

Fig. 4 has been prepared to show some points connected with the time separation of the strokes in those discharges which consisted of a long series. Each stroke is represented by a point and each series is connected by lines. The number of the flash is attached to the series. The ordinates of the diagram give the time interval separating a stroke from its predecessor, hence the series begins at the left with the second stroke and ends on the right with the last stroke. In spite of a tendency towards uniformity shown by flashes 56 and 38, the majority of the time intervals in any given series are very irregularly distributed.

---

* 'Ann. Rep. Smithsonian Inst.,' p. 119, 1905.

Flashes 23 and 24, which lasted for 0·89 and 0·93 seconds respectively, were the longest in total duration.

TABLE II—TIME INTERVALS BETWEEN STROKES

Interval ranging between	Number recorded
0·001 secs and 0·002 secs	1
0·002 ,, 0·005 ,,	6
0·005 ,, 0·01 ,,	9
0·01 ,, 0·03 ,,	28
0·03 ,, 0·05 ,,	22
0·05 ,, 0·07 ,,	11
0·07 ,, 0·09 ,,	15
0·09 ,, 0·11 ,,	4
0·11 ,, 0·15 ,,	6
0·15 ,, 0·23 ,,	1
0·23 ,, 0·32 ,,	3
0·32 ,, 0·42 ,,	1
0·42 ,, 0·53 ,,	2

(*c*) *Relation between Order of Stroke and Intensity*—Out of 33 discharges with more than one stroke the first stroke of the series was the most intense in 29 (88%). In two the second stroke, in one the third, and in one both second and third strokes were more intense than the first.

(*d*) *Relation between Intensity and the Time Interval before Occurrence of a Stroke*—In fig. 4, those strokes which have an intensity greater than the first stroke (not included in the diagram) are marked with a large double circle. It will be seen that they are in general strokes which occurred after a longer interval of waiting than the average. In three out of the six cases this interval exceeds 0·1 sec and in five cases it is greater than 0·06 sec.

That stroke in each series which has the greatest intensity after the first is marked in fig. 4 with a black circle, and here also there is a pronounced tendency for the greatest intensity to be associated with the longest time interval. The third stroke of flash 75 (75*c* of fig. 3, Plate 16), occurring after an interval of 0·48 sec, is an illustration of these findings.

(*e*) ***Branching of the Discharge***—Out of 57 first strokes examined for evidence of branches, only three were unbranched, and these were on photographs classified as fair in one case and poor in the two others.

A total of 40 first strokes on photographs classified as good and fair showed 198 branches, an average of five branches per stroke. No cases of upward branching were observed on any of these discharges between cloud and ground.

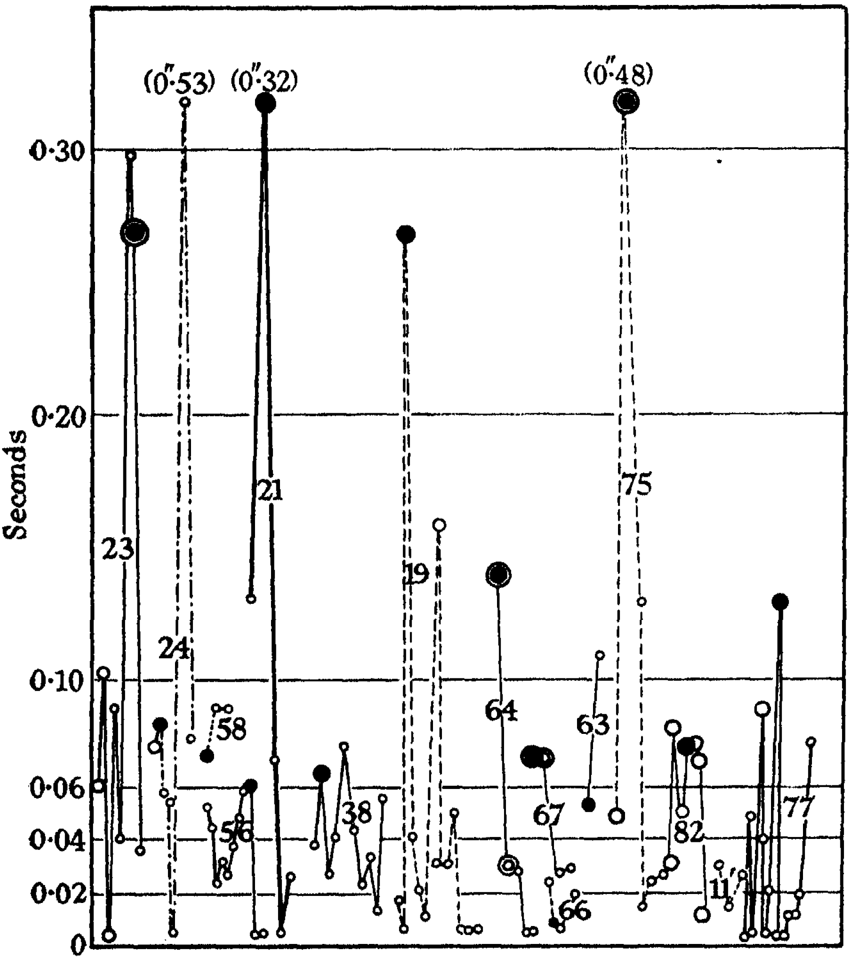


FIG. 4—◉ intensity > 1st stroke; ● maximum intensity of series after first stroke.

While 54/57, or 95% of these first strokes showed branches, it was observed on only 14 out of 33 subsequent strokes (42%), the average number of branches being one per stroke for these 14. The branches on subsequent strokes were generally much weaker in intensity than those on first strokes.

Of a total of 18 branches on subsequent strokes, all except two doubtful cases followed the course of branches already traced out by the first strokes. In some instances, however, they developed new forks at their ends.

The difference* between first and subsequent strokes in respect of branching affords a useful means of locating first strokes where no slow camera is employed.

Brief mention may here be made of the fact that close to their lower ends these flashes generally show one or more nearly vertical branches which appear to reach to the ground, as if the leader broke up into a number of forks, the first of which to hit the earth formed the trunk of the return stroke. Such "*root branching*," as it may be termed, occurs at a distance of from 20 to 150 metres from the ground, and the forks at their lower ends are 20 to 80 metres apart.

(*f*) *Permanence of Form of Discharge Channel*—In the great majority of cases in which a discharge consists of a series of strokes the main channel keeps its form unaltered, the same twists and bends being observed on all the strokes. In 4 out of 87 discharges studied, however, it was found that changes occurred in the channel, the top portion altering in three and the bottom portion to ground in one. When such a change in the channel occurred it persisted for the strokes which followed. These exceptions will be further discussed in a later publication.

5—LEADER STROKES

(*a*) *Occurrence of the Leader-Return Stroke Combination*—In order to determine whether the leader-main stroke sequence was the most frequent, if not the invariable, mode of development in these discharges, the data were analysed as shown in Table III, which indicates the numbers of first and of subsequent strokes showing leaders.

Of the "good" pictures, 93% show leaders, while of "good" and "fair" combined the fraction is 73%. Including pictures classed as "poor" we find 94 leaders to 158 strokes.

Where no leader portion of a stroke was observable, the subsequent main return portion always travelled upwards from ground to cloud as usual, and this fact, coupled with the figures in Table III, leads us to conclude that the leader was too faint to be recorded. All the return portions of the strokes photographed by us have been found to travel upwards, and of the 33 complete flashes analysed in Table III, 28 show leaders on one or more of their strokes.

---

* This was, of course, noticed by earlier workers with slow cameras, *e.g.*, Walter, *loc. cit.*

In these thunderstorms, therefore, the leader-return stroke sequence was in the great majority of cases, if not always, the characteristic feature of discharges to ground.

(*b*) *First Stroke Leaders*—The leaders to first strokes can usually be most easily recognized on the slow-moving camera, where they have the dart-like appearance shown in fig. 2. On the fast Boys camera they are difficult to see unless the discharge is near, for the series of steps in their progress may occupy almost the full circumference of the circular track, much of which is taken up with other strokes and general fogging. On the best pictures we can see the whole course of these stepped leaders, but on many others only a portion is visible, depending on the amount of background present. Here recourse is had to the slow-moving camera to show that the leader track was actually present all the way down from cloud to ground.

TABLE III—OCCURRENCE OF LEADERS

Photography	First Strokes		Subsequent Strokes		All Strokes	
	Number	Leaders	Number	Leaders	Number	Leaders
Good	15	15	26	23	41	38
Fair	18	9	41	26	59	35
Totals	33	24	67	49	100	73

Table IV summarizes the manner in which information of this kind has been collected. In the first column we have described the photographs from the point of view of the possibility of detecting the rather faint steps on the fast camera record.

TABLE IV—LEADERS TO FIRST STROKES

Photography	Number of first strokes	Leaders observed on			
		Slow and fast cameras	Slow only	Fast only	None
Good	15	7	5	3	0
Fair	18	3	1	5	9
Poor	19	0	3	0	16

No case has been observed in which the leader to a first stroke has been other than stepped when examined with the fast camera.

(*c*) *Subsequent Stroke Leaders*—The leaders to subsequent strokes are generally of a continuous dart-like character, as described in I, p. 662, but in at least three discharges we have definite evidence of the "dart-stepped" form, in which the dart changes at its lowest and slowest end to a stepped leader. The steps in such cases, as shown in Table V, are very short, about 10 metres long, and the pauses between them about 8 microseconds, so that it is not easy to detect them except under good conditions. It is possible that the effect occurs in other cases as well and that what has been described as the dart-leader process is actually only a special case of the stepped leader to the first stroke. In Table V the length found to be stepped, shown in column 5, is the length nearest the ground.

TABLE V—DART-STEPPED LEADERS TO SUBSEQUENT STROKES

Flash	Stroke	Velocity (cm/sec 10^7)	Total length km	Length stepped km	Step length metres	Pause time μ-sec
67	*b*	1·9	2·02	0·3	9·0	7·4
64	*b*	1·2	2·3	0·7	10·0	9·0
75	*d*	1·0	3·0	0·7	7·4	7·4

The detailed discussion of stepped leaders, in particular of those to first strokes, is resumed in § 8.

6—THE VELOCITIES OF LEADERS

(*a*) *Effective Velocity of Stepped Leaders*—The stepped leaders to first strokes (§§ 3 and 8) proceed in a series of jerks to ground, and consequently their *effective velocity*, *v*, is considerably less than the velocity V with which the streamer producing each fresh step advances. The effective velocity is obtained by dividing the total path length, L, by the time occupied over the path, T, and the values found for this quantity are given in the distribution Table VI. The table contains corresponding data for a number of stepped discharges from the cloud which do not reach the ground (*air-discharges*), and for the stepped portions of dart-stepped leaders to subsequent strokes.

The mean value of *v* for leaders to ground from first strokes is $3 \cdot 8 \times 10^7$ cm/sec, but about half the values listed lie within the range $1 \cdot 0$ to $3 \cdot 0 \times 10^7$ cm/sec. The above are two-dimensional path velocities and should be increased by 30% to obtain the actual velocities in three dimensions (I, p. 660). The values for the five air discharges and for dart-stepped leaders are higher than for ordinary stepped leaders.

The most probable length of time taken by a stepped leader to traverse 1 km of air would, from Table VI, be 0·007 sec, the range being from 0·01 to 0·0008 sec. For three first-leader processes between a cloud at an undetermined height and the ground, Walter* gave durations of 0·02, 0·009, and 0·007 sec.

TABLE VI—EFFECTIVE VELOCITIES OF STEPPED LEADERS

Velocity range (cm/sec × 10^7)		First stroke leaders to ground	Air discharges	Stepped parts of dart-stepped leaders
From	To			
1·0	2·0	9	0	0
2·0	3·0	3	0	0
3·0	5·0	3	0	0
5·0	7·0	3	1	0
7·0	9·0	3	3	0
9·0	11·0	1	0	1
11·0	13·0	1	1	0
13·0	15·0	0	0	1
18·7		0	0	1

(*b*) *Velocities of Dart Leaders to Subsequent Strokes*—The distribution of the average velocities of dart leaders is shown in Table VII. As before, these are two-dimensional track velocities. Data for three air-discharges of the dart variety (§ 10) are included in the table.

TABLE VII—DISTRIBUTION OF DART-LEADER VELOCITIES

Velocity range in cm/sec × 10^8		
From	To	Dart leaders
1·0	3·0	17
3·0	5·0	8
5·0	7·0	6 + 1*
7·0	9·0	5 + 1*
9·0	11·0	5
11·0	13·0	2
13·0	15·0	6 + 1*
15·0	17·0	1
17·0	19·0	2
19·0	21·0	2
21·0	23·0	1

* Cloud-air discharge.

The mean velocity of these 55 leaders is 5·5 × 10^8 cm/sec and the value indicated as most probable is 2·0 × 10^8 cm/sec. It has been shown (I,

---

* Walter, *loc. cit.*, pp. 21 and 30.

p. 660) that the velocity often varies along the path, becoming less as the leader approaches the ground. This statement has been confirmed with the new material, and no instance of the dart-leader velocity increasing at the lower end has yet been observed. It is noteworthy that three of the slow leaders in the first line of Table VII are definitely stepped in their later stages (§ 5*c*).

(*c*) *Correlation between Dart-Leader Velocities and other Factors*—In order to determine whether any relation exists between dart-leader velocity and other qualities of the flash as a whole or of the particular stroke itself, we have searched for correlation with the number of strokes making up the discharge, with the time interval which elapsed before the stroke occurred, and with the intensity of the main part of the stroke as compared with that of the first stroke of the series. Information on these points is contained in Table VIII for slow leaders (velocities from 1·0 to $3 \cdot 0 \times 10^8$ cm/sec) and for fast ones (from $15 \cdot 0$ to $21 \cdot 5 \times 10^8$ cm/sec).

In view of the irregular manner in which the average time interval varies from flash to flash, § 4*d* and fig. 4, it is not to be expected that Table VIII would show any very close relation between leader velocity and time interval in general. It does, however, exhibit a tendency for the slower leaders to be associated with the longer intervals. Flash 75 offers a good example, giving a fast leader when the interval before the (fifth) stroke is 0·014 sec, and a slow one when that before the (third) stroke is 0·48 sec. The same point is brought out more clearly when the intensity of the main return part of the stroke is considered. This is given as a rough estimate, taking that of the first stroke as unity, in column 5 of Table VIII, and it is evident that slow leaders are associated with strokes which carry more energy than usual. Indeed, all the strokes in which the main part has a greater intensity than the first main stroke are included in the slow leader portion of Table VIII. It has already been seen that high intensity is in general associated with an unusually long interval before the occurrence of the stroke, § 4 (*d*). The slow leader stroke of 75*c* exhibits an intensity ratio of 2·0/1 in its subsequent return portion, while the fast one 75*e* shows 0·4/1.

The increase of the intensity of the dart leaders themselves with velocity, which was noted in the previous paper (I, p. 657), is abundantly confirmed in the present work and will be further examined by a photometric method.

7—(*a*) *The Downward Branching of Leader Strokes*—Branching during the downward movement of leaders to first strokes was observed in 16 of the 33 first strokes listed in the first and second lines of Table III, The total number of leader branches was 52. On strokes other than the first the feebleness and rarity of leader branching makes it more difficult to detect, but it was observed in a number of these cases also. Examples of first-leader branching are shown in figs. 7, 8, 10, 11, 13, Plates 17, 19, and 20. There can thus be no doubt that the characteristic downward branching of a lightning flash between cloud and ground arises in the first instance during the progress of the downward-moving leader-stroke.*

TABLE VIII

	Stroke	Strokes in flash	Time interval from previous stroke sec	Intensity*	Velocity cm/sec
Fast leaders	66*b*	5	0·015	0·3	$21 \cdot 5 \times 10^8$
	23*b*	9	0·06	0·4	$17 \cdot 1 \times 10^8$
	23*c*	9	0·12	0·4	$20 \cdot 2 \times 10^8$
	19*c*	14	0·005	0·8	$19 \cdot 3 \times 10^8$
	75*e*	7	0·014	0·4	$15 \cdot 0 \times 10^8$
Slow leaders	67*b*	5	0·07	3·0	$1 \cdot 9 \times 10^8$
	67*c*	5	0·07	1·5	$1 \cdot 7 \times 10^8$
	75*c*	7	0·48	2·0	$1 \cdot 8 \times 10^8$
	11*b*	4	†	5·0	$1 \cdot 7 \times 10^8$
	82*e*	8	0·08	1·2	$2 \cdot 0 \times 10^8$
	64*b*	6	0·14	3·0	$2 \cdot 8 \times 10^8$
	67*e*	5	0·03	1·5	†

* First stroke = 1·0.
† Not observable.

The detailed development of leaders and their branches can be observed on the fast-moving Boys camera, which shows the separate steps in the progress of first leaders, and also on the slow-moving single-lens camera, which does not resolve the steps. From either of these, but most easily from the latter, time-tables can be drawn up giving the progress of events in the leader. For, since the comparatively small distortion of the return stroke channel is known from the Boys camera record, the true leader distortion can be determined by a comparison of leader and return stroke pictures obtained by the slow single-lens camera.

(*b*) *Time-tables for Leaders*—Time-tables of this kind are shown in figs. 5 and 6 for strokes 28*a*, 69*a*, and 70*a*, of figs. 7 and 8, Plate 17. The leader time-scales, marked in light lettering, are in *milliseconds* measured from the *tops* of the channels of 28 and 69 respectively. The return stroke times, marked in heavy lettering, are measured in *microseconds* from the *bottoms* of 28*a* and 69*a*. In each the direction of movement of the camera lens is indicated by an arrow, the recorded leader track by the fainter line, and the return stroke track by the heavier line.

---

* *See also* Walter, 'Ann. Physik,' vol. 52, p. 421 (1935), who points out that this deduction was clearly made by him from slow camera observations in 1902.

All three flashes give several leader branches and side-forks, which are later more strongly illuminated by the return strokes.

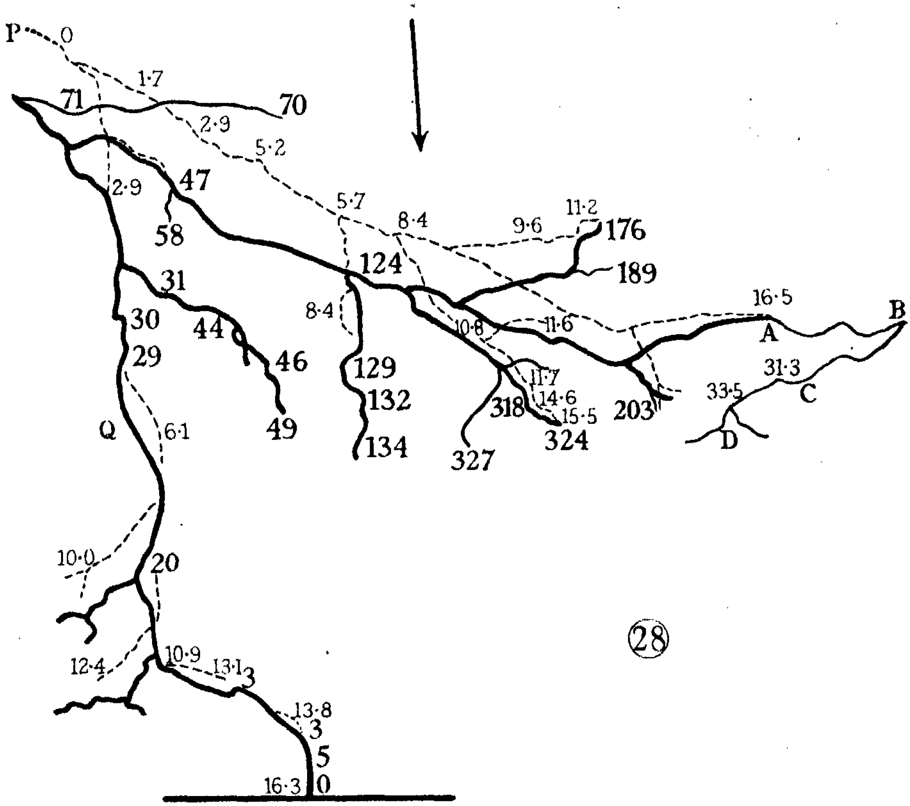


Fig. 5.

Flash 28, shown in fig. 7, Plate 17, on all three cameras, gives on the Boys instrument a stepped leader, the details of which are too faint for reproduction. The fixed camera caught only a portion of the main trunk but shows most of the long horizontal branch. The leader time-table is obtained from the slow camera record in the lower right-hand corner. All the leader branches of the stroke are marked out on this record *above* the brighter traces made by the return portion since the lens was moving *downwards* in the direction of the arrow. The leader branch first formed

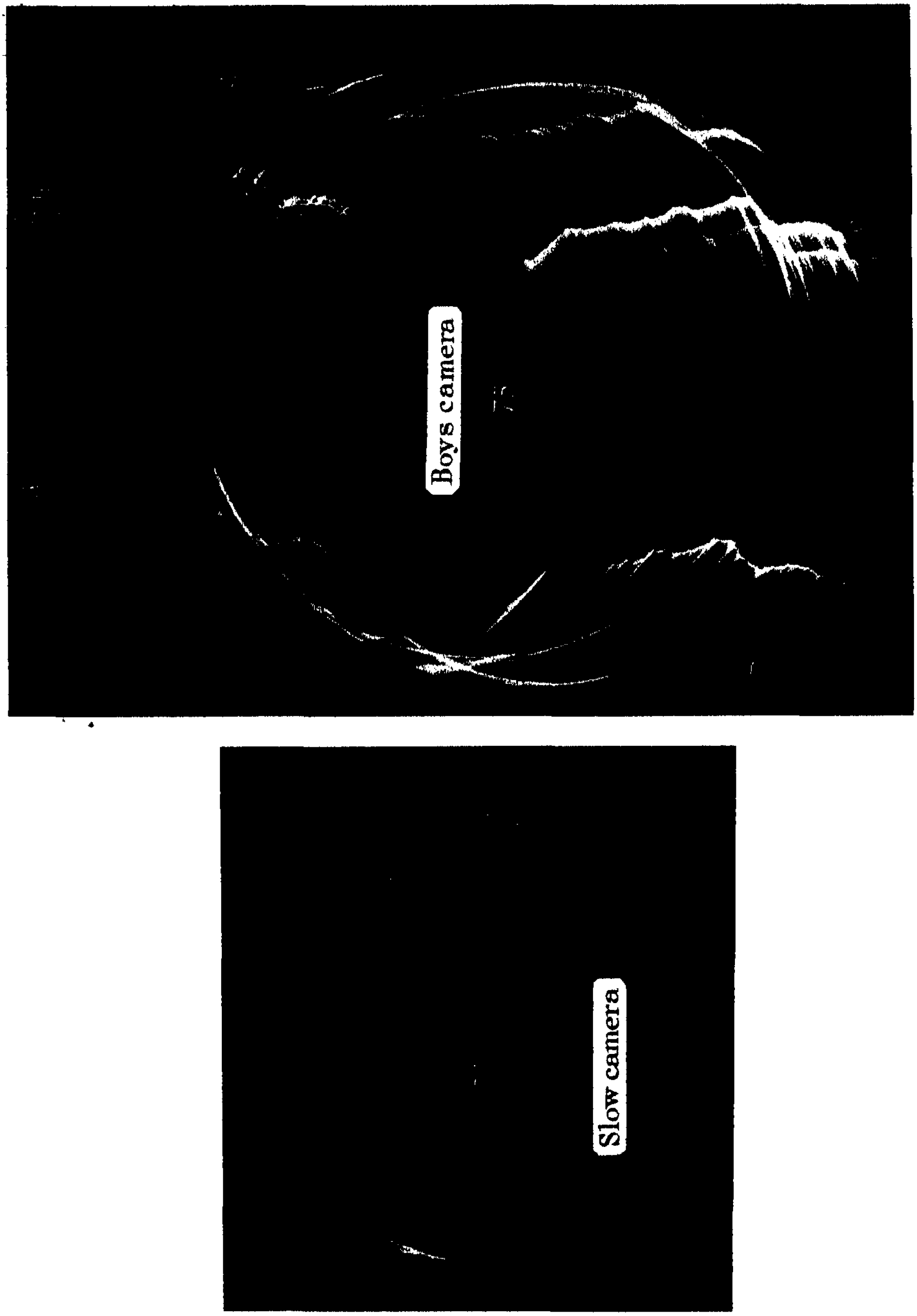


FIG. 3

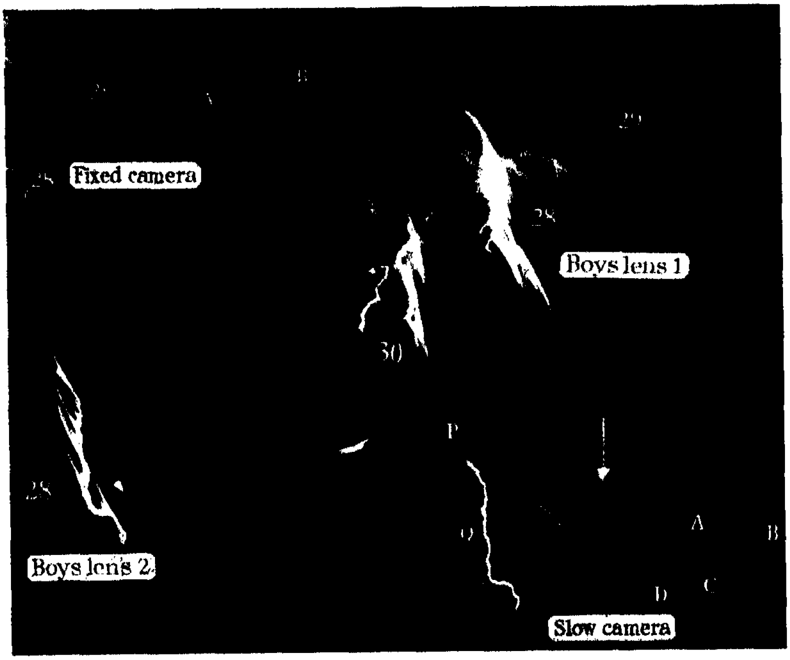

B
29
Fixed camera
28
Boys lens 1
30
P
28
Q
A
B
C
D
Boys lens 2
Slow camera

FIG. 7.

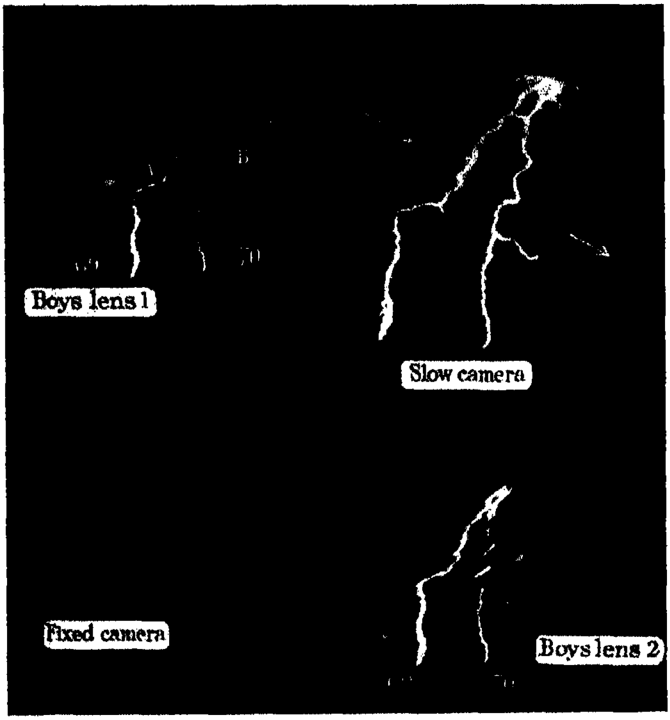

B
30
Boys lens 1
Slow camera
Fixed camera
Boys lens 2

FIG. 8.

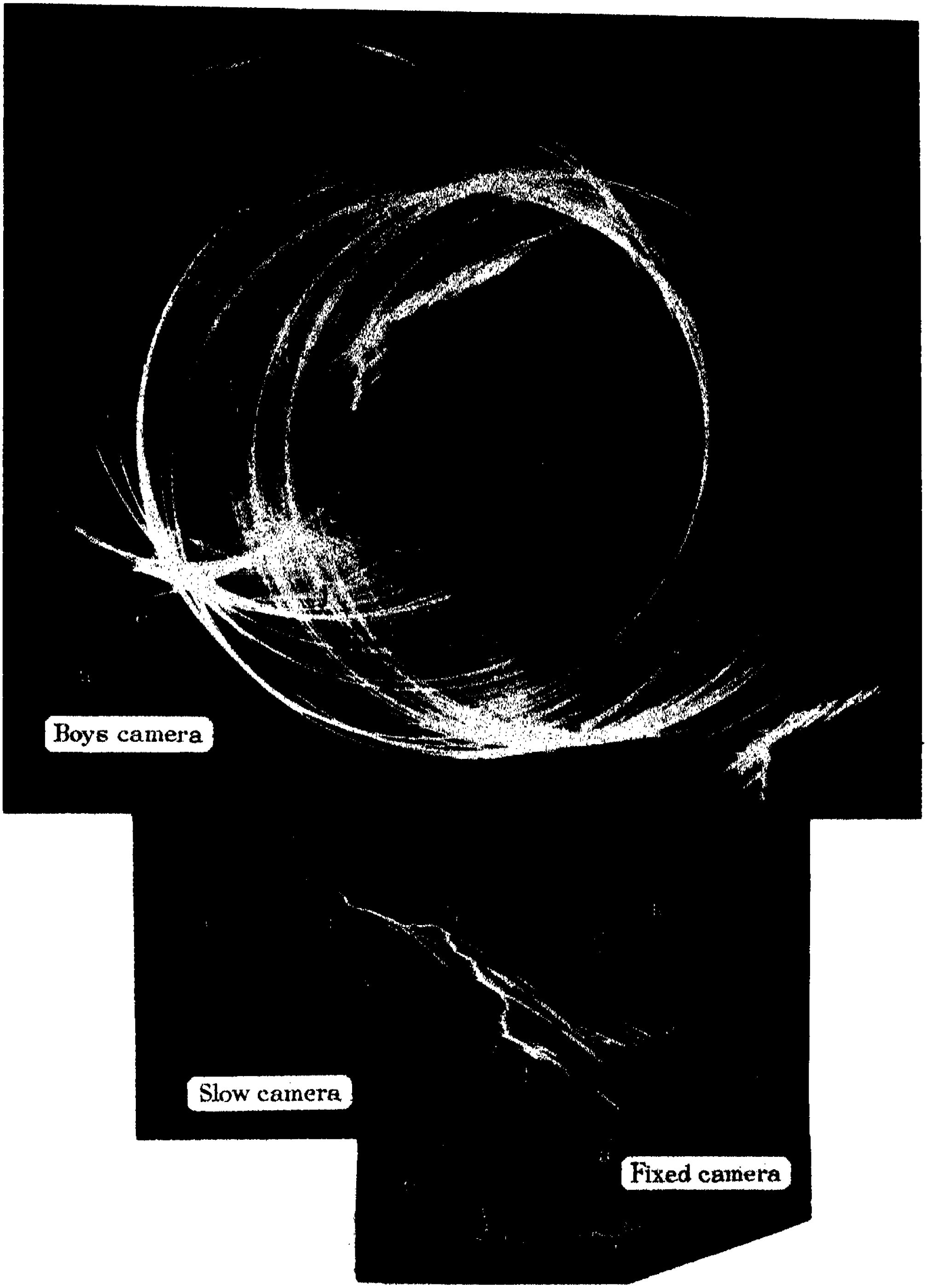


Fig. 9.

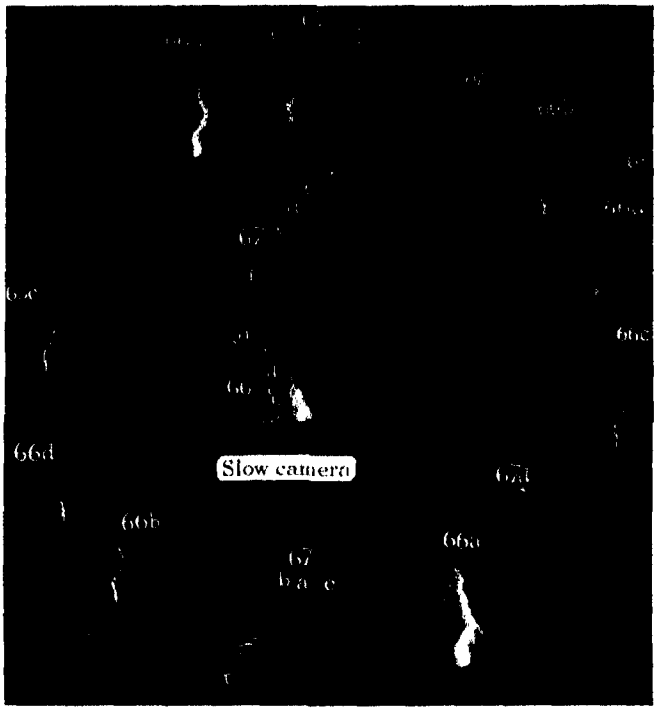


FIG. 10.

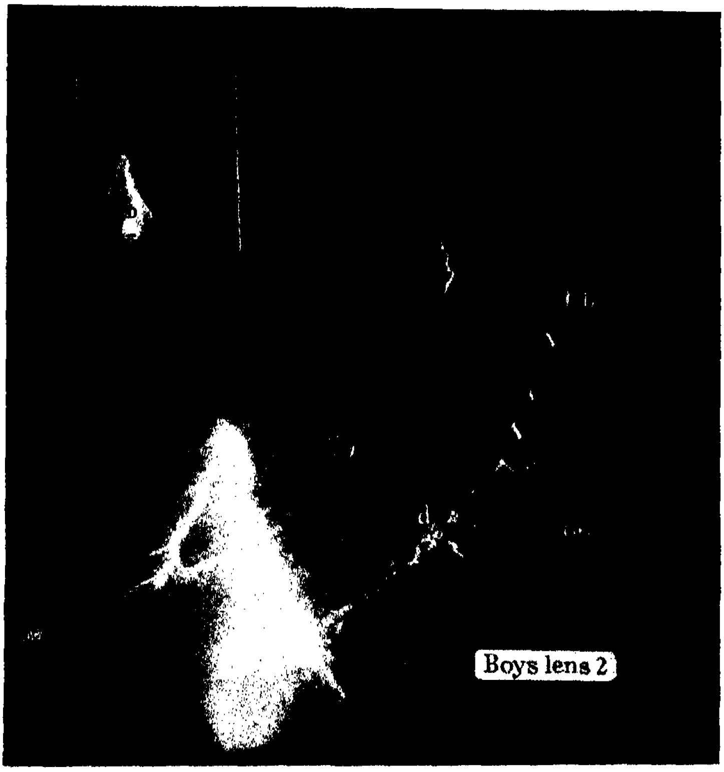


FIG. 11.

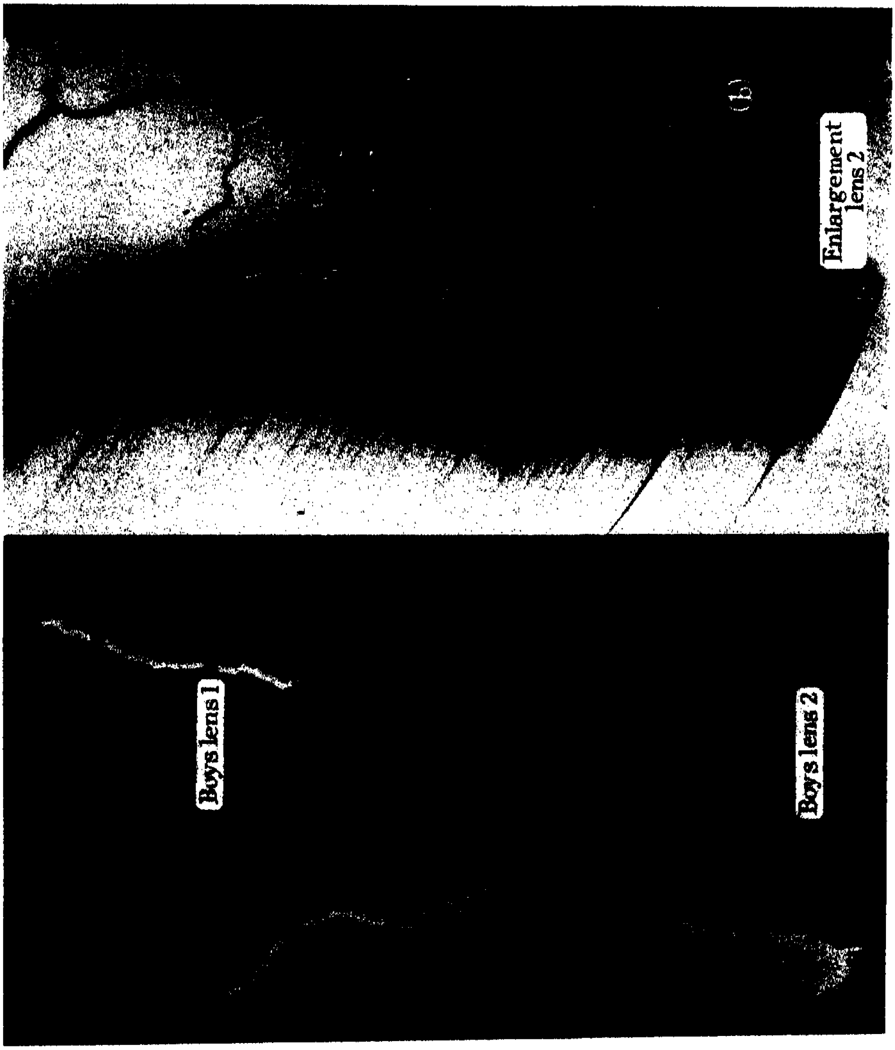
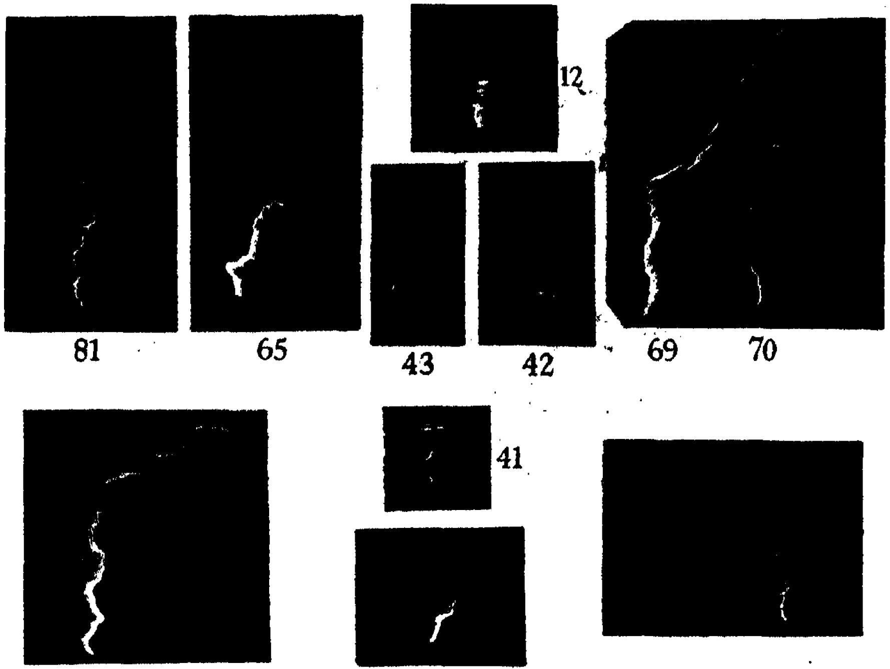


FIG. 13.



81 65 12 43 42 69 70

33 41 16 87

Fig. 17.

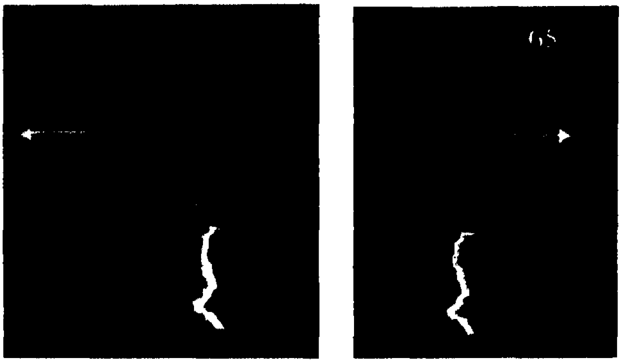


Fig. 20.

shows a very pronounced continuation or drag of luminosity, indicating that the agency pouring energy outwards along the branch was not entirely concentrated at the advancing tip. This drag is evidently due to the luminosity of the fainter streamers referred to in § 3, fig. 2.

The effective velocity of 28*a* along its main channel was $3 \cdot 8 \times 10^7$ cm/sec from P to Q, and $1 \cdot 7 \times 10^7$ cm/sec from Q to the ground, while the long branch was formed at $3 \cdot 0 \times 10^7$ cm/sec.

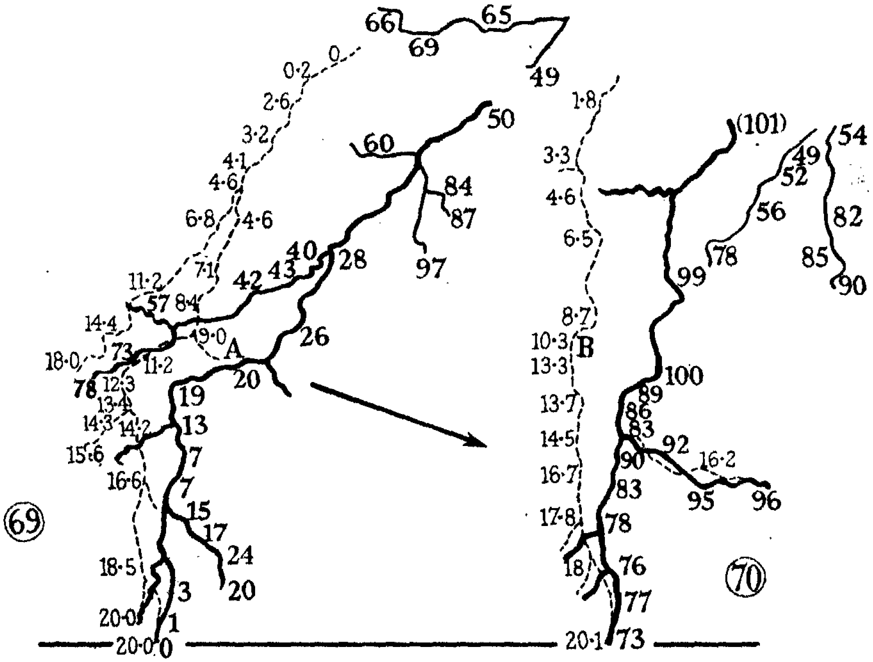


Fig. 6.

Flashes 69 and 70, whose complete record is given in fig. 8, Plate 17, with a time-table in fig. 6, also had stepped leaders too faint to show on the reproduction. They seem to have been twin strokes from the same region of the cloud, perhaps branching from a single trunk higher up. The slow-camera record of their development gives the time-table of fig. 6, in which the leader times are in milliseconds measured from the moment of appearance of the top of flash 69. The top portion of flash 70 appears 0·0018 sec later, but this flash, by taking a shorter path to ground, is only 73 microseconds later than flash 69 in hitting the earth. The path-lengths and effective velocities of the leaders to 69 and 70 are 2·6 and 2·2 km, $1 \cdot 30 \times 10^7$ and $1 \cdot 22 \times 10^7$ cm/sec respectively. Some values of the effective velocities along different parts of their courses are given below.

Flash 69—	Top to A, uniform	$1\cdot3 \times 10^7$ cm/sec.
	A to bottom, uniform	$1\cdot2 \times 10^7$ cm/sec.
	Along branch, practically uniform	$1\cdot3 \times 10^7$ cm/sec.
Flash 70—	Top to B, uniform	$1\cdot2 \times 10^7$ cm/sec.
	B to bottom, uniform	$1\cdot2 \times 10^7$ cm/sec.
	Along branch, uniform	$1\cdot0 \times 10^7$ cm/sec.

The uniformity in the values of v for these cases is not a general rule. Cases occur in which the stepped leader has a much higher value of v at the start than later on, and others in which it speeds up at its lower end. Only two general statements can be made: the effective leader velocity is found to be low when executing a path which is nearly horizontal in the record, and to be high during the last hundred metres before striking the earth. The first effect may be due to an underestimation of the actual track-length when the track appears horizontal, but the second is evidently due to the increase in field as the leader-tip approaches the ground. Flash 69, for example, develops a leader velocity of $3\cdot0 \times 10^7$ cm/sec in the last 70 metres of its path.

8—The Nature of Stepped Leaders and Air Discharges

(*a*) *Some Special Cases*—In amplification of the descriptions given in § 3, three of the twenty-four cases of stepped leaders and air discharges, § 10, will be described in some detail.

Flash 36, fig. 9, Plate 18, was an air discharge which passed horizontally from the cloud-base into the air and then turned upwards, following a path indicated in the figure by the letters *xfsbjce*, placed at easily recognizable twists and branches. The whole of the discharge, which was 10 km in length, was not recorded by the fixed camera (lowest picture). The slow camera (centre picture) shows that the time occupied in passage from right to left was 0·007 seconds between *s* and *e* (8 km). The Boys double lens camera (upper picture) resolved the discharge into a series of streamers with bright tips, the average step length being 67 metres and the average time interval between steps 74 microseconds. Between *s* and *e* there were approximately 110 steps, and the fast-moving lenses each made nearly half a revolution during the progress of the discharge. The points marked on the lower pictures are reproduced in the top one at the places where the streamer series first reached them. It will be seen that the streamers increase in brightness and decrease in width as they travel towards the left from their starting-point on the right. Since this is an air discharge (§ 10) there is no return main stroke at its end, but it shows a slow " recoil " effect (§ 9) at the end of branch *a*, marked with the letter R on the slow camera record. Here the discharge returns over a distance of 2·6 km with a velocity of $5 \cdot 9 \times 10^7$ cm/sec.

Flash 35, the return portion of which was recorded on these pictures, was a discharge to the ground from the same region of the cloud, and will not be further discussed here.

In fig. 10, Plate 19, are reproduced the records of the four strokes of flash 66, which passed between cloud and ground at a distance of 12 km from the cameras. The leader process for 66*a* was so slow that it was only satisfactorily recorded on one of the Boys camera lenses, and this record is reproduced from an intensified positive copy in fig. 11, Plate 19, with a sketch diagram to scale in fig. 12. The clarity of the record of the main stroke has been unavoidably spoilt in intensification and the original picture is therefore shown in fig. 11 as an inset.

In studying figs. 11 and 12 it must be remembered that the lens motion was from right to left, and that the time is measured by displacement along the circular arc traversed by the ray through the centre of the lens, so that the illumination of a point at the top of the channel, for example, is to be found at various times along the arc *a*A. The leader began at *a* and ended below *e* at a time 0·0221 seconds later. The trunk of the leader channel followed the same course as the subsequent rapid main portion of the discharge (FDCBA) but in the opposite direction, and this course, owing to frequent pauses, was recorded by the moving lens along *abcde*. A branch developed from the leader at its starting-point *a* and extended in the same channel as that subsequently, brightly and rapidly, followed by the main stroke from A. Similarly at *b*, *c*, and *d*, the leader steps separate out to form the branches B, C, D. The last stages of the development of the leader trunk are visible at *e*, where it reaches to within 51 metres of the ground in its penultimate step.

The leader reaches the ends of the upper branches A and B some 800 μ-sec before the main stroke arrives there, and so appears to cease to develop these branches some time before reaching the ground. Branches C and D, however, were still being pushed outwards by the leader at the time it reached the ground, and the leader steps for these two were caught up by the rapid upward-moving main stroke.

In this picture the streamers extending from the cloud-base downwards are just visible in certain of the stronger steps and then towards their lower ends only. Higher up they become wider and too faint to be distinguished from the general background.

Comparison of the two Boys camera records of the return main stroke shows that the total time taken by it to pass from F to A was 55·0 microseconds.

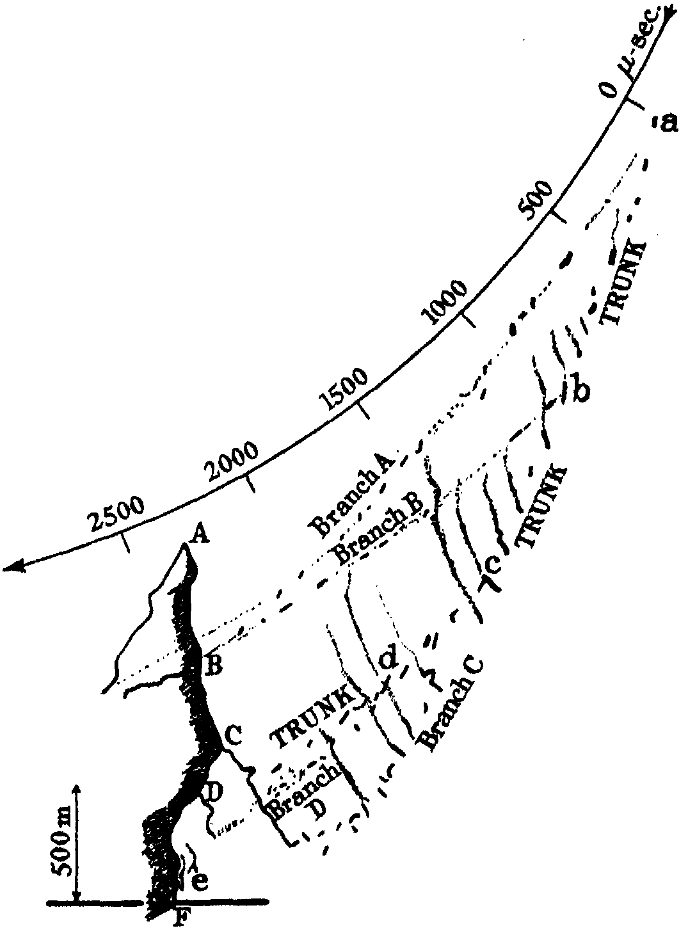


FIG. 12.

Flash 76, fig. 13, Plate 20, and fig. 14, took place between cloud and ground at a distance of 2·0 ± 0·5 km, so close that the upper part of the stroke was slightly out of focus for one of the Boys camera lenses and off the film for the other, fig. 13*a*. The flash, which consisted of a single stroke, was recorded on process film, the fine grain of which enabled a valuable record of the details of the leader to be obtained. These details are very clear in the lower half of the picture, which has been enlarged and intensified for fig. 13 (*b*) and is sketched to scale in fig. 14.

The leader trunk follows the course *abcdefg* corresponding to

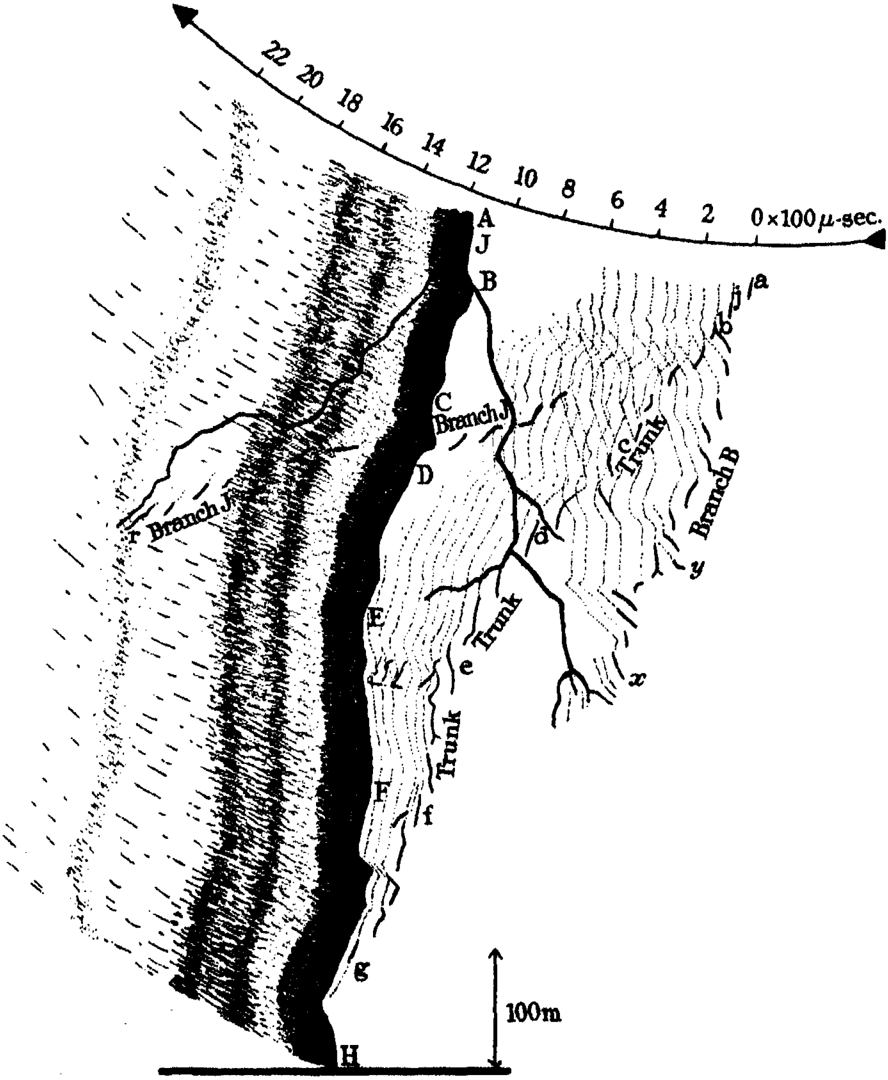


FIG. 14.

ABCDEFH on the rapid return stroke. The lens movement in figs. 13 and 14 was along a circular arc from right to left. The leader branched at *j* to form branch J, and at *b* to form branch B, the subsequent steps of which form a series arranged along *bx*, with two side-forks developing from them. The details of these side-forks, one of which starts at *y*, have been omitted to avoid confusion. For the same reason only a certain number of the bright steps passing from *j* to *r* to form branch J have been inserted in fig. 14, and their faint streamers have been omitted altogether.

The last step *r* of branch J traces out the final section of this branch and the rapid return stroke goes no farther than this step. The interval between *r* and the arrival of the return stroke is 18 microseconds, equal to the time interval determined by the Boys method between the start of the return stroke at the base and the end of branch J. Thus leader development of branch J ceased at the moment the leader trunk hit the ground.

The details of the leader steps at the end of branch B are more difficult to follow, and it would seem as if this branch faltered in its progress before the trunk arrived at the ground.

The streamers which end in these prominent bright tips are, as fig. 13, Plate 20, shows, narrower and brighter at their lower ends, and they become difficult to follow as one passes upwards from their tips. This effect has not been shown in fig. 14. Normally, the luminosity of the bright tips appears to cease abruptly, but the penultimate step *g* in the leader trunk is exceptional in showing a faint continuation some 30 metres long towards the ground. A similar effect, but to a much smaller extent, is shown by a few other steps higher up.

The lengths of the bright tips shown in fig. 13 vary from 36 to 15 metres, with an average value of 24·6 metres. The variation in length must, of course, be due to some extent to the effect of motion in the line of sight. The time intervals between the ends of steps and the beginnings of their successors vary from 96 to 16 μ/sec, with a mean of 52 microseconds. As the streamers approach the ground the step-lengths increase and the intervals decrease. Between *a* and *e* their average values are 21·5 metres and 63 microseconds, and between *e* and *g* they are 27·0 metres and 42 microseconds. The last interval of all, at *g*, is only 16 microseconds, and the last step, as already mentioned, is unusually long. The effective velocity of the leader process is $3 \cdot 4 \times 10^7$ cm/sec from *a* to *e*, and $6 \cdot 4 \times 10^7$ cm/sec from *e* to *g*.

Other details of these steps are included in the next two sections.

(*b*) *Summary of Details of Leader Steps*—The following summary contains a list of the main features of the leader steps, quantitative study of some of them being left to the next section. The statements made are based on the full material available to us,* and reference is made where necessary to figs. 12 and 14 as illustrations.

(1) Each bright step appears as the termination of a fainter streamer extending the whole way down from the starting-point of the discharge, such streamers increasing in brightness and decreasing in width as their ends are approached.

(2) A bright step generally starts a little way back on part of the track formed by the previous step, so that the portion of the step which is entirely new is about 90% of the whole, fig. 14, step after *e*.

(3) The form of the channel blazed by a new step is not always followed in detail by the streamers of subsequent steps, the tendency being for the smaller twists and kinks to be smoothed out. In a few cases the form of the subsequent streamer is changed so completely as to suggest that a fork previously existed, the fainter branch of which was adopted as the channel in later stages, fig. 14, second step after *e*.

(4) A beaded effect at the beginning and end of the bright steps is frequently found, which in view of its position cannot be due to motion in the line of sight.

(5) Branching occurs as a division of one step into a fork, fig. 12, *c*, and fig. 14, *b* and *y*, the subsequent development of a branch taking place in steps simultaneous with those of the trunk. Short branches and branches near the top of the channel, however, may cease to develop some time before the main trunk steps have reached the ground. The start of a branch never occurs elsewhere than at the beginning of a bright step.

(6) Each fresh step takes in general a different direction to its predecessor, and it is unusual to find a sharp bend in the course of a single step. The tortuous form of a lightning discharge channel thus originates in the different directions taken by successive steps.

(7) The streamers and their tips are generally much brighter in their later stages, when they are approaching the ground, than earlier. It is not always possible to detect even the tips in the upper portion of the leader. This increase in brightness is associated with a considerable increase in effective velocity near to the ground.

(8) The brightness of the streamers and their tips is found to increase with increase in the effective velocity, v (§ 6*a*), of the leader process. The three cases discussed at length in § 7*a* are discharges in which v is higher than the average.

(9) Some hesitancy is occasionally shown in the development of both branch and trunk steps. Thus the branch which started at *e* in fig. 14

---

* 24 cases in all.

missed three steps and then re-blazed the first branch step twice. At the same spot the trunk development missed a step and made up for it by an exceptionally long jump the following step. Similar cases occur at *y* and at *f* in fig. 14, and at the step two below *c* on branch C of fig. 12.

(*c*) *Velocity of Leader Streamers*—To determine the velocity of the streamers which make up the leader steps one can either use the Boys method of comparing one rotating lens record with the other, or compare the track of the streamer, as recorded by one lens, with that of the return stroke, whose distortion is easily found by the Boys method. Difficulties, however, arise from the limited lengths of streamer tracks available for measurement and the lack of sharpness in their leading edges. Such comparisons have not so far yielded any evidence of distortion of the streamer track as a result of lens motion, and it is concluded from several measurements that the streamer velocity exceeds 5×10^9 cm/sec.

Definite evidence that the streamers travel downwards is, however, afforded by the broadening of the upper part of their tracks. If we take the duration of luminosity, t, at a point distant d from the tip as being equal to the time taken for the streamer to travel this distance we again obtain, as an upper limit for the velocity, values of d/t of the order of 5×10^9 cm/sec. Unfortunately, the method is too rough to indicate whether the streamer velocity varies at different points along the streamer or in different streamers.

The tips of the streamers, which from their sudden cessation of luminosity might be expected to show a decrease in velocity, exhibit no distortion due to lens motion and have velocities which are certainly greater than 10^9 cm/sec.

(*d*) *Relations between Step-length, Step-interval, and Effective Velocity*—It is convenient to denote by $\bar{l}$ the average length, along the two-dimensional record, of the leader steps, by $\bar{t}$ the average time interval between steps, so that the quantity $\bar{l}/\bar{t}$ is an average velocity. If the time occupied in tracing the step-length $\bar{l}$ is small compared with $\bar{t}$, this velocity should be the same as the effective velocity v (L/T) with which the stepped leader advances. Comparison of v with $\bar{l}/\bar{t}$ shows them to be identical within the accuracy of measurement, as indeed would follow from the high value of the streamer velocity along $\bar{l}$ found in the preceding section.

Values of $\bar{l}$ and $\bar{t}$ have been obtained with sufficient accuracy from 11 first-stroke leaders, three cloud-air discharges, and three dart-stepped leaders. The remaining seven cases of the stepped process are too poor to give reliable data. The results are plotted in fig. 15, which shows the dart-stepped cases with crosses and the remainder with circles. The black circles indicate cases where the photography was good and the measurements could be made with more accuracy than for the open circle cases. It will be seen that while $\bar{l}$ varies from 10 to 206 metres, a ratio of 1/20, $\bar{t}$ for first strokes only alters from 31 to 91 microseconds, a ratio of 1/3. The larger step-lengths are associated with the longer time intervals. The same type of variation in the course of a single leader has already been noted for flash 76.

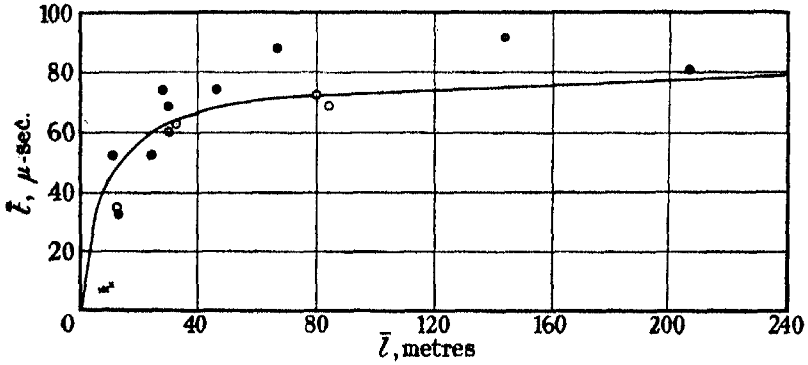


FIG. 15.

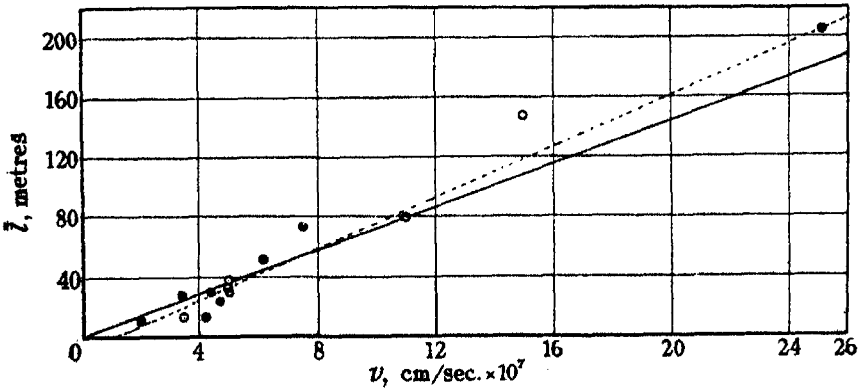


FIG. 16.

To a rough first approximation $\bar{t}$ might be considered as constant, and hence it is that when in fig. 16 the values of $\bar{l}$ are plotted against those of v, or $\bar{l}/\bar{t}$, the points lie fairly well on a straight line through the origin. A better fit, however, is obtained in fig. 16 with the dotted line which cuts the horizontal axis at $v = 1{\cdot}0 \times 10^7$ cm/sec, a value for which, as will be shown later, there is theoretical justification. From this line it follows that the relation between $\bar{l}$ and $\bar{t}$ is $t = 80\,\bar{l}/(\bar{l} + 800)$, where the units are microseconds and centimetres. The curve of fig. 15 has been drawn from this equation.

The corresponding relation between $\bar{t}$ and v is $\bar{t} = 80\,(v - 10^7)/v$, where the units are microseconds and cm/sec.

9—RETURN PORTIONS OF STROKES

(*a*) *Decrease of Intensity after Passing Branch Points*—It was noted in I (p. 669) that the main trunk illumination in the upward-moving return stroke was sometimes found to decrease markedly in intensity at each branch point from the bottom upwards. After the last such branch-point the intense " bright " luminosity was found to be replaced by a weaker " glow " luminosity. The material now available offers numerous examples of this effect, which has been observed in 29 out of 39 of the " good " and " fair " photographs, while in four out of the 10 cases in which it was not observed this is to be attributed to the presence of a prominent branch at the top end of the channel.* A selection of examples of this effect is shown in fig. 17, Plate 21, and by flashes 66, 67, and 75 of figs. 10 and 3, Plates 19 and 16. It is impossible to avoid the conclusion that " the activity which makes the channel brightly luminous is engaged primarily in pouring energy into the branches " (I, p. 669).

Unbranched subsequent strokes show an effect of the same character, for the intensity of the bright luminosity decreases in general from the base upwards. No case has yet been observed in which the trunk luminosity was greatest at the cloud end of the discharge.

The average distance over which the bright luminosity of the return portion of first strokes extends is 0·72 of the whole track. Values as low as 0·2 have been found, and a more detailed photometric study is planned.

(*b*) *Velocities*—The velocities of 32 return or main strokes were given in the previous paper (I, p. 663), and the distribution of these, together with 16 new ones, is shown in the first column of fig. 18. The range of variation is from $2 \cdot 0$ to $14 \cdot 4 \times 10^9$ cm/sec, and the mean $5 \cdot 2 \times 10^9$ cm/sec, while the value indicated as most frequent is $3 \cdot 5 \times 10^9$ cm/sec.

In view of the facts given in § 9*a* concerning the change from bright to glow luminosity at the last branch from the ground, the velocities in the bright portions, column 2, and the glow portions, column 3, are of interest. It will be seen that the former are in general about twice as great as the latter, and that the bright luminosity is associated with velocities of the order of $1 \cdot 0 \times 10^{10}$ cm/sec. The fourth column of fig. 18 gives return stroke velocities along five particularly good branches, each belonging to a different stroke. These are of the same order as the velocities in column 1.

---

* *E.g.*, flash 28 of fig. 7, Plate 17.

(*c*) *Time-tables*—The order of events in the return stroke of flash 65, which was single and 2·5 km long, is shown in fig. 19. The times are given in microseconds from the start of the return portion from the base.

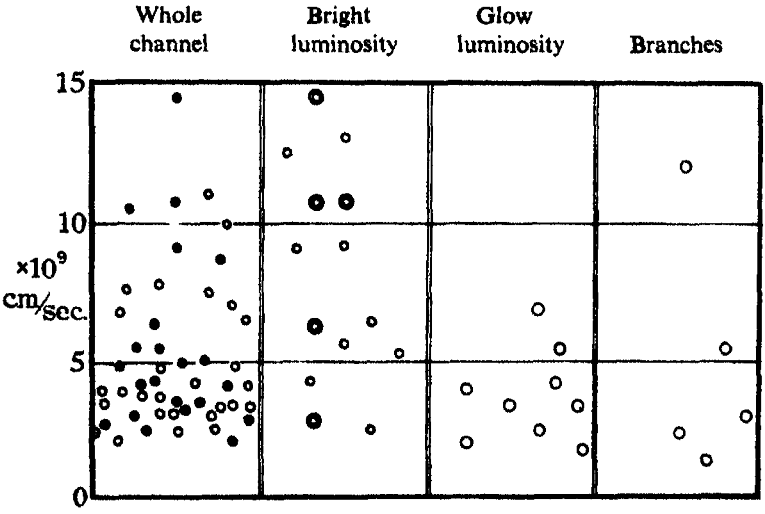


FIG. 18.

This stroke, shown in fig. 20, Plate 21, gave a particularly good edge for measurement,* and the velocities at different stages are of interest. These are listed below.

	cm/sec
Base to Branch 2†	$16 \cdot 1 \times 10^9$
Branch 2 to Branch 3	$21 \cdot 0 \times 10^9$
Branch 3 to Branch 5	$9 \cdot 7 \times 10^9$
Branch 5 to top	$5 \cdot 5 \times 10^9$

Owing to the short lengths involved, the values given are perhaps only accurate to within 20%, but they bring out very clearly the point made above that the brighter portions near the base are travelling the quickest.

---

* The error of measurement is less than 1 μ-sec, except at the top of the main trunk and along parts of the branches, where it is 2 μ-secs.

† Numbered from base upwards.

In view of the accuracy with which the times of fig. 19 can be found it is interesting to notice that there is evidence of a definite pause, lasting 5 ± 1 microseconds, at the point P, the start of the third and longest branch. Here the bright return stroke appears to run along the branch for about 100 metres before proceeding upwards. After P the bright luminosity is much reduced, and the stroke obviously concentrates its main effort on the portion up to and including the third branch.

The records obtained for flash 28 are shown in fig. 7, Plate 17. The leader times, obtained from the slow-moving camera record in the lower right-hand corner, have been discussed in § 7. The return stroke time-table obtained from the Boys camera record is a very good one and shown in fig. 5, by the heavy figures, where the times are measured in microseconds from the base upwards. The average velocity along the trunk is 10·6 × 10^9 cm/sec; along the branch it is 1·8 × 10^9 cm/sec, the return stroke travelling outwards more slowly than usual.

After the return stroke has reached the point A and caught up with the branch leader, it extends a further 1·9 km to the point B. Then there occurs a curious "recoil" effect which has already been noted for a branch of flash 36, and which causes the luminosity to return slowly back along B C D with a velocity of 1·0 × 10^7 cm/sec. The original extension from A to B is made at a velocity of 1·32 × 10^7 cm/sec.

These movements, which can only be detected on the slow-camera record, are probably stepped in nature.

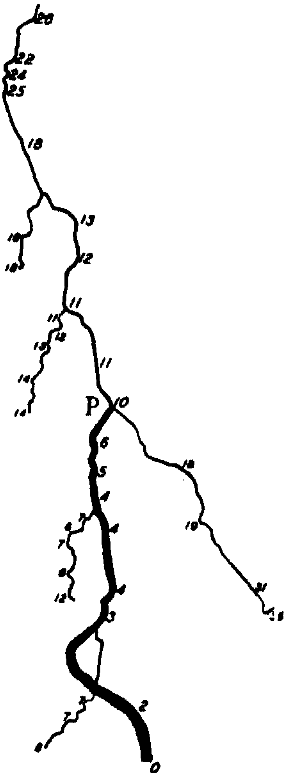


FIG. 19.

65

Flash 30, which occurred so far to the left in fig. 7, Plate 17, that it was only recorded by one Boys camera lens, offers no information as to its nature.

Flashes 69 and 70, figs. 6 and 8, Plate 17, whose leaders have been described in § 7*b*, give return stroke time-tables shown in microseconds by the heavy figures of fig. 6, in which the zero of time is the start of the return stroke of 69*a*. The velocities over different portions of the two discharge channels are listed below.

		cm/sec
Flash 69—	Average track velocity	$5 \cdot 0 \times 10^{9}$
	Bottom to Branch 3[*]	$5 \cdot 2 \times 10^{9}$
	Branch 3 to Branch 5	$8 \cdot 0 \times 10^{9}$
	Branch 5 to top	$3 \cdot 3 \times 10^{9}$
	Along Branch 5	$2 \cdot 5 \times 10^{9}$
Flash 70—	Average velocity	$7 \cdot 5 \times 10^{9}$
	Bottom to Branch 3	$9 \cdot 2 \times 10^{9}$
	Branch 3 to top	$6 \cdot 7 \times 10^{9}$
	Along Branch 3	$12 \cdot 0 \times 10^{9}$[†]

The falling off of velocity with distance from the base, and after heavy branching, is more marked in the case of flash 69 than in flash 70.

All the time-tables discussed above exhibit the regular nature of branch illumination during the return stroke ; each branch is traversed at a time close to, if not identical with, the arrival of the return stroke at the branching-point, and the direction of movement is outwards from the trunk in all cases except the two short " recoil " lengths referred to above.

10—AIR DISCHARGES

Discharges which do not reach the ground offer an interesting problem. They appear to be of two distinct types. In the first, best illustrated by flash 36, fig. 9, Plate 18, the discharge consists of a succession of streamers exactly like the stepped leaders, advancing in a series of steps from the cloud into the air. The effective velocities of advance of four out of the seven cases observed are given in Table VI of § 6*a*, and their step-lengths and step-intervals are included in figs. 15 and 16.

The second type of cloud-air discharge is one in which the mode of progression is that of the dart type of leader. The velocities determined for three of the seven cases found are shown in Table VII.

The stepped type of discharge is generally associated with the heavier cloud-air flashes, and it is possible that the dart type follows upon a stepped discharge too weak to detect. In neither type is there evidence of a return main stroke from the end of the channel, but a " recoil " effect has been noted for one case at least (R, fig. 9).

We have no information as to the nature of the discharges which take place within the cloud itself, but a visual observation of an unusual nature may be recorded. A particularly active and extraordinarily transparent

[*] Numbered from base upwards.

[†] Approximate.

thunderstorm observed at a distance of 60 km gave numerous discharges within the cloud, all of which branched upwards and appeared to travel slowly upwards from the base. No downward discharges were observed, and the impression left on the observers (B. S. and D. M.) was that they were witnessing an upwardly directed discharge exactly similar to the downward-stepped leader of a flash to ground, and without a return main stroke.

11—General Discussion

It has seemed best to present the facts collected in the foregoing sections with little if any commentary of a theoretical nature. The detailed theoretical discussion will shortly be attempted in a further publication, but it is desirable to terminate the present account with a short résumé of those points which seem to emerge without difficulty from the data.

It is probable that practically every one of the many strokes comprised in the flashes studied was double and consisted of a leader and a return portion, § 5*a*. The first-leader process involves the formation of a conducting channel along which flows electric charge from the cloud region tapped. Its progress is slow compared with that of the other and later processes involved in the discharge, and it always occurs in a stepwise manner, § 5*b*. In the course of its downward movement the leader branches outwards into regions where a local concentration of space-charge may be presumed. Evidence of this space-charge is suggested by the recoil effect observed at some branches, § 9*c*, and by the root-branching effect, § 4*d*, in the last 100 metres of its path.

The advancing leader with its branches constitutes an actual downward movement of part of the cloud-charge, which is distributed over the whole conducting system, and in particular over the branches, at the moment the leader hits the ground. The leader process thus effectively lowers a portion of the cloud-charge into the air. When this charged system is placed in good conducting connexion with the earth, a rapid flow of charge takes place, on a scale and at a rate such as is exhibited by the main return stage of the stroke. The frequently observed concentration of the energy of the return stroke upon the lower two-thirds of the channel and its branches, § 9*a*, as well as its decrease in velocity as it passes the branches, § 9*c*, offers evidence that the leader process often lowers into the air the greater part of the cloud-charge tapped. An exactly similar effect is observed in subsequent strokes; the downward-moving leader again charges up the channel and the return stroke intensity and duration indicate by the manner in which they fall off with height above the ground that the leader process has drained the cloud region it taps of most of the available charge.

The very great intensity of the return-stroke process is evidently due to the presence at the end of the charged conducting leader channel of a good conducting earth, with the induced charge present at or near the point struck. For discharges into the air only, the absence of such a terminating conductor accounts for the absence of any return stroke, except in the form of recoil strokes of small intensity and long duration, § 9*c*.

It is to be expected that the rather striking difference between dart leaders to subsequent strokes and stepped leaders to first strokes is one of degree rather than of kind, the dart leader form arising when the leader disturbance passes along a previously ionized channel. This expectation is confirmed by the occurrence of the stepped form at the lower ends of very slow dart leaders, § 5*c*. The velocity of an electrical disturbance along an ionized channel must depend largely upon the conductivity per unit length, if this is not of the order of the conductivity of a metal. A freshly ionized channel would therefore be expected to transmit the dart-leader disturbance with a higher velocity than one in which cooling and recombination have been taking place for some time. The data concerning the variation of dart-leader velocity with the time interval which has elapsed since the previous stroke, § 6*c*, are in satisfactory accord with this view. They show, indeed, by the fact that the slowest leaders are associated with the oldest channels, although followed by the most intense return strokes, that leader velocity is far more dependent upon channel conductivity than upon the charge and potential difference driving the leader forward.

When the interval since the last occurrence of a stroke is excessively long and the ionization in the channel has therefore fallen to a lower value than usual, it is found that the leader disturbance is unable to continue in a dart-like manner and becomes stepped. The stepped leader is thus indicated as the only possible method of progress in a channel which is either not sufficiently ionized or has not previously been ionized at all.

When the age of the channel is small and the conductivity big the dart leader velocity is found to rise as high as 2×10^9 cm/sec. For channels which carry the streamers making up the separate steps of first leaders, the previous ionization is not particularly intense, but it is very recent, and such streamers thus move in a continuous manner with a velocity estimated to be of the order of 5×10^9 cm/sec. For the return stroke a continuous development at a still higher velocity is observed.

Without entering into a full discussion on this point it is necessary to note that the control exercised by the pre-existing conductivity of the channel upon the velocity of advance is not so direct as it would be in a poorly conducting metal. The rate of progress of a discharge depends ultimately upon the strength of the electric field at its tip, and this in turn upon the potential drop along, and so the resistance of, the newly blazed portion of the channel. A smaller pre-existing degree of ionization at the tip means fewer ionizing collisions and consequently less ionization in the new portion of the channel behind the tip. This in its turn must involve a higher resistance between cloud and tip, a larger potential drop, a smaller field at the tip and thus a smaller velocity of advance. In conformity with this view it is found, § 6*c*, that the light emitted at the tips of dart leaders, which indicates the amount of ionization produced there, decreases with decreasing velocity. The same principles can be used to provide a quantitative explanation of the stepped-leader advance.

This work has been carried out under the direction of the Lightning Research Committee of the South African Institute of Electrical Engineers. In addition to the acknowledgments made in the text, we desire to thank Sir Charles Boys and Mr. Bernard Price for their encouragement and interest. Acknowledgment is gratefully made by one of us (B. S.) to the Carnegie Corporation of New York and the South African Research Grant Board for the grant of a Travelling Fellowship, and to Professors H. Paine and A. Ogg for providing generous laboratory facilities in Johannesburg and Cape Town.

12—SUMMARY

A general account, based on the study of 95 flashes with the Boys and other cameras, is given of the mode of development of the lightning discharge. A statistical study is made of the distribution of the strokes of a discharge in respect of number, time-interval, intensity, and manner of branching. It is shown that the leader-return stroke sequence is present in almost all the cases examined. Leaders to first strokes are stepped, those to subsequent strokes generally dart-like. In certain cases of very slow dart leaders these change to the stepped form at their lower ends. The effective velocities of stepped leaders range from $1\cdot0 \times 10^7$ to $2\cdot0 \times 10^8$ cm/sec, the velocities of dart leaders from $1\cdot0 \times 10^8$ to $2\cdot3 \times 10^9$ cm/sec, and of main return strokes from $2\cdot0 \times 10^9$ to $1\cdot4 \times 10^{10}$ cm/sec. The most frequent values for the three types of development are $1\cdot5 \times 10^7$, 2×10^8, $3\cdot5 \times 10^9$ cm/sec respectively.

It is shown that slower leader velocities and higher intensities of return strokes are associated with longer time-intervals between strokes and their

predecessors, and hence that the degree of pre-existing ionization in the channel governs the velocity of the dart leader.

The downward-branching of lightning and its characteristic zig-zag form are shown to arise during the stepped-leader process before the first return stroke.

An account, based upon 24 cases examined, is given of the nature of the stepped leader. The velocity of the streamers making up the steps is of the order of 5 × 10^9 cm/sec. The range of variation of the time intervals between steps is small (31–91 μ-sec) compared with that of the step-lengths themselves (10–206 metres).

A decrease in the intensity of the return portion of a stroke in passing branch-points is found to be general, and it is deduced that the leader process lowers into the air a very considerable fraction of the cloud charge tapped.

Discharges into the air exhibit no return stroke but sometimes show a slow recoil effect from their ends.

The Relativistic Self-Consistent Field

By Bertha Swirles, Ph.D., University of Manchester

(*Communicated by D. R. Hartree, F.R.S.—Received June* 12, 1935)

1—The method of the self-consistent field for determining the wave functions and energy levels of an atom with many electrons was developed by Hartree,[†] and later derived from a variation principle and modified to take account of exchange and of Pauli's exclusion principle by Slater[‡] and Fock.[§] No attempt was made to consider relativity effects, and the use of " spin " wave functions was purely formal. Since, in the solution of Dirac's equation for a hydrogen-like atom of nuclear charge Z, the difference of the radial wave functions from the solutions of Schrödinger's equation depends on the ratio Z/137, it appears that for heavy atoms the relativity correction will be of importance; in fact, it may in some cases be of more importance as a modification of Hartree's original self-consistent field equation than "exchange" effects. The relativistic self-consistent field equation neglecting "exchange" terms can be formed from Dirac's equation by a method completely analogous to Hartree's original derivation of the non-relativistic self-consistent field equation from Schrödinger's equation. Here we are concerned with including both relativity and "exchange" effects and we show how Slater's variational method may be extended for this purpose.

[†] ' Proc. Camb. Phil. Soc.,' vol. 24, p. 89 (1928).

[‡] ' Phys. Rev.,' vol. 34, p. 1293 (1929).

[§] ' Z. Physik,' vol. 61, p. 126; vol. 62, p. 793 (1930); vol. 81, p. 195 (1933); V. Fock and M. Petrashen, ' Phys. Z. Sowjet.,' vol. 6, p. 368 (1934).

A difficulty arises in considering the relativistic theory of any problem concerning more than one electron since the correct wave equation for such a system is not known. Formulae have been given for the interaction energy of two electrons, taking account of magnetic interactions and retardation, by Gaunt,† Breit‡, and others. Since, however, none of these is to be regarded as exact, in the present paper the crude electrostatic expression for the potential energy will be used. The neglect of the magnetic interactions is not likely to lead to any great error for an atom consisting mainly of closed groups, since the magnetic field of a closed group vanishes. Also, since the self-consistent field type of approximation is concerned with the interaction of *average* distributions of electrons in one-electron wave functions, it seems probable that retardation does not play an important part. These effects are in any case likely to be of less importance than the improvement in the grouping of the wave functions which arises from using a wave equation which involves the spins implicitly.

It is first shown that Dirac's equations for a single electron can be derived from a variation principle and this is extended to the case of many electrons when the magnetic interactions and retardation are ignored. The wave functions for individual electrons are assumed to be quantities with four components of the types found for a single electron in a central field§ and the anti-symmetrical wave function for the whole atom to be a "determinant" built up from these in the same way as the non-relativistic wave function for a whole atom is built up from one-electron wave functions. It is then possible to obtain an expression for the total energy of the atom in terms of radial functions only. This necessitates the construction of tables analogous to those given by Slater.‖ Then by a variation method differential equations for these radial wave functions can be obtained. As an example, we derive the equations for the two different types of $2p$ wave function for copper.

† 'Phil. Trans.,' A, vol. 228, p. 151 (1929).

‡ 'Phys. Rev.,' vol. 34, p. 553 (1929).

§ Darwin, 'Proc. Roy. Soc.,' A, vol. 118, p. 654 (1928).

‖ 'Phys. Rev.,' vol. 34, p. 1312 (1929).

2—Dirac's expression† in atomic units for the hamiltonian H for a single electron moving in a field of force of scalar potential V and zero vector potential is the matrix

$$\mathrm{H} = -\mathrm{VU} + c\boldsymbol{\alpha} \cdot \mathbf{p} + c^2\beta, \tag{1}$$

where c is equal to 137 in atomic units, $\mathbf{p} = -i\nabla$, U is the unit matrix with four rows and columns, and the components of $\boldsymbol{\alpha}$ $(\alpha_1, \alpha_2, \alpha_3)$ and β are given by

$$\alpha_1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad \alpha_2 = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix},$$

$$\alpha_3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{2}$$

We form the volume integral

$$\mathrm{E} = \int \psi_\rho^* \mathrm{H}_{\rho\sigma} \psi_\sigma \, d\tau \tag{3}$$

taken over all space. The suffixes ρ, σ go from 1 to 4, and we use the summation convention throughout. We shall show that Dirac's equations can be obtained by requiring that this integral shall be stationary for arbitrary independent variations of the four components of ψ, subject only to a normalizing condition. We first show that the integral (3) is real for any ψ obeying a given boundary condition, not only for solutions of Dirac's equations. For

$$\mathrm{E} = \int \psi_\rho^* (-\mathrm{VU}_{\rho\sigma} - ic\boldsymbol{\alpha}_{\rho\sigma} \cdot \nabla + c^2\beta_{\rho\sigma}) \psi_\sigma \, d\tau, \tag{4}$$

and since U and β are real, the conjugate complex quantity E* is

$$\mathrm{E}^* = \int \psi_\rho (-\mathrm{VU}_{\rho\sigma} + ic\,\boldsymbol{\alpha}_{\rho\sigma}^* \cdot \nabla + c^2\beta_{\rho\sigma}) \psi_\sigma^* \, d\tau. \tag{5}$$

Now,

$$\mathrm{U}_{\rho\sigma} = \mathrm{U}_{\sigma\rho}, \quad \beta_{\rho\sigma} = \beta_{\sigma\rho} \quad \text{and} \quad \boldsymbol{\alpha}_{\rho\sigma}^* = \boldsymbol{\alpha}_{\sigma\rho}.$$

---

† Pauli, 'Handb. der Physik,' vol. 24, p. 233.

Therefore†

$$\mathrm{E}^* = \int \psi_\sigma^* \left(-\mathrm{VU}_{\sigma\rho} - ic\boldsymbol{\alpha}_{\sigma\rho} \cdot \nabla + c^2\beta_{\sigma\rho}\right) \psi_\rho \, d\tau + ic \int (\psi_\rho \boldsymbol{\alpha}_{\rho\sigma}^* \psi_\sigma^*) \cdot \mathbf{dS}, \quad (6)$$

where the surface integral is taken over the sphere at infinity and therefore vanishes if all the components ψ_ρ are $o(1/r)$ for large r. The volume integral in (6) is the same as (4) with a change of the dummy suffixes. It follows that, subject to the condition for the behaviour at infinity, $\mathrm{E} = \mathrm{E}^*$, so that for any such ψ E defined by (3) is real.

We normalize the ψ's to unity, thus

$$\int \psi_\rho^* \psi_\rho \, d\tau = 1, \quad (7)$$

and consider the result of putting

$$\delta \mathrm{E} = 0 \quad (8)$$

subject to

$$\delta \int \psi_\rho^* \psi_\rho \, d\tau = 0.$$

This is in complete analogy to the procedure in the non-relativistic case, since, if the ψ_ρ are wave functions, (3) is the expression for the total energy of the system while $\psi_\rho^* \psi_\rho$ is the charge density. We have then

$$\int (\delta\psi_\rho^* \mathrm{H}_{\rho\sigma} \psi_\sigma + \psi_\sigma^* \mathrm{H}_{\sigma\rho} \delta\psi_\rho) \, d\tau - \varepsilon \int (\delta\psi_\rho^* \psi_\rho + \psi_\rho^* \delta\psi_\rho) \, d\tau = 0, \quad (9)$$

where ε is real since it is equal to a stationary value of

$$\frac{\int \psi_\rho^* \mathrm{H}_{\rho\sigma} \psi_\sigma \, d\tau}{\int \psi_\rho^* \psi_\rho \, d\tau},$$

which is real as proved above.

Now,

$$\int \psi_\sigma^* \mathrm{H}_{\sigma\rho} \delta\psi_\rho \, d\tau = \int \delta\psi_\rho \mathrm{H}_{\rho\sigma}^* \psi_\sigma^* \, d\tau - i \int (\psi_\sigma^* \boldsymbol{\alpha}_{\sigma\rho} \, \delta\psi_\rho) \cdot \mathbf{dS}, \quad (10)$$

---

† Here and in equation (10) we apply the divergence theorem to $\int \operatorname{div} (\psi_\rho \boldsymbol{\alpha}_{\rho\sigma}^* \psi_\sigma) \, d\tau$ instead of to $\int \operatorname{div} (\psi \nabla \psi) \, d\tau$ as in the non-relativistic case.

and the surface integral over the sphere at infinity vanishes under the given conditions. (9) may therefore be written in the form

$$\int [R(\delta\psi_\rho)(\mathrm{H}_{\rho\sigma}\psi_\sigma + \mathrm{H}_{\rho\sigma}{}^*\psi_\sigma{}^* - \varepsilon\psi_\rho - \varepsilon\psi_\rho{}^*) \\ + I(\delta\psi_\rho{}^*)(\mathrm{H}_{\rho\sigma}{}^*\psi_\sigma{}^* - \mathrm{H}_{\rho\sigma}\psi_\sigma + \varepsilon\psi_\rho - \varepsilon\psi_\rho{}^*)]\, d\tau = 0. \tag{11}$$

If we are considering a general variation of ψ we must regard each of its components as capable of variation independently of the others and also must contemplate independent variations of $R(\delta\psi_\rho)$ and $I(\delta\psi_\rho)$, the real and imaginary parts of $\delta\psi_\rho$. The coefficient of $R(\delta\psi_\rho)$ is purely real and that of $I(\delta\psi_\rho)$ purely imaginary. We have therefore

$$\mathrm{H}_{\rho\sigma}\psi_\sigma + \mathrm{H}_{\rho\sigma}{}^*\psi_\sigma{}^* - \varepsilon\psi_\rho - \varepsilon\psi_\rho{}^* = 0 \tag{12}$$

and

$$\mathrm{H}_{\sigma\rho}{}^*\psi_\sigma{}^* - \mathrm{H}_{\sigma\rho}\psi_\sigma + \varepsilon\psi_\rho - \varepsilon\psi_\rho{}^* = 0. \tag{13}$$

By adding and subtracting these equations we have

$$\mathrm{H}_{\rho\sigma}\psi_\sigma - \varepsilon\psi_\rho = 0 \tag{14}$$

$$\mathrm{H}_{\sigma\rho}{}^*\psi_\sigma{}^* - \varepsilon\psi_\rho{}^* = 0, \tag{15}$$

which are Dirac's equations for ψ and ψ^* and are written more fully as

$$(-\mathrm{V}\mathrm{U}_{\rho\sigma} - ic\,\boldsymbol{\alpha}_{\rho\sigma}\,.\,\nabla + c^2\beta_{\rho\sigma} - \varepsilon\mathrm{U}_{\rho\sigma})\,\psi_\sigma = 0 \tag{16}$$

and

$$-\mathrm{V}\psi_\sigma{}^*\mathrm{U}_{\sigma\rho} + ic\,\nabla\psi_\sigma{}^*\,.\,\boldsymbol{\alpha}_{\sigma\rho} + c^2\psi_\sigma{}^*\beta_{\sigma\rho} - \varepsilon\psi_\sigma{}^*\mathrm{U}_{\sigma\rho} = 0. \tag{17}$$

For an electron in a field of potential V, a function of r only, equations (16), written out in full, become

$$\left.\begin{aligned} i\left(-\frac{\varepsilon+\mathrm{V}}{c}+c\right)\psi_1 + \left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right)\psi_4 + \frac{\partial}{\partial z}\psi_3 &= 0 \\ i\left(-\frac{\varepsilon+\mathrm{V}}{c}+c\right)\psi_2 + \left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)\psi_3 - \frac{\partial}{\partial z}\psi_4 &= 0 \\ -i\left(\frac{\varepsilon+\mathrm{V}}{c}+c\right)\psi_3 + \left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right)\psi_2 + \frac{\partial}{\partial z}\psi_1 &= 0 \\ -i\left(\frac{\varepsilon+\mathrm{V}}{c}+c\right)\psi_4 + \left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)\psi_1 - \frac{\partial}{\partial z}\psi_2 &= 0 \end{aligned}\right\} \tag{18}$$

Darwin† has shown that two types of solution in polar co-ordinates may be obtained for these equations, the one corresponding in the usual notation to $(l, j = l + \frac{1}{2}, m = u + \frac{1}{2})$ and given by†

---

† 'Proc. Roy. Soc.,' A, vol. 118, p. 654 (1928). The spherical harmonics used here are the conventional ones, not the modified ones used by Darwin. Differences in sign between equations (19) and Darwin's (7.1) are due to a different formulation of Dirac's hamiltonian. The notation used here is that used by Pauli ("Handbuch der Physik," vol. 24).

$$\left.\begin{aligned}
\psi_1 &= -\,i\mathrm{F}_l\,(l - u + 1)\,\mathrm{P}_{l+1}^{u}(\mu)\,e^{iu\phi} \\
\psi_2 &= -\,i\mathrm{F}_l\,\mathrm{P}_{l+1}^{u+1}(\mu)\,e^{i(u+1)\phi} \\
\psi_3 &= \quad \mathrm{G}_l\,(l + u + 1)\,\mathrm{P}_{l}^{u}(\mu)\,e^{i(u+1)\phi} \\
\psi_4 &= -\,\mathrm{G}_l\,\mathrm{P}_{l}^{u+1}(\mu)\,e^{i(u+1)\phi}
\end{aligned}\right\}\; 0 \leqslant u \leqslant l \qquad (19\text{A})$$

$$\left.\begin{aligned}
\psi_1 &= -\,i\mathrm{F}_l\,\mathrm{P}_{l+1}^{-u}(\mu)\,e^{iu\phi} \\
\psi_2 &= i\mathrm{F}_l\,(l + u + 2)\,\mathrm{P}_{l+1}^{-(u+1)}(\mu)\,e^{i(u+1)\phi} \\
\psi_3 &= \mathrm{G}_l\,\mathrm{P}_{l}^{-u}(\mu)\,e^{iu\phi} \\
\psi_4 &= \mathrm{G}_l\,(l - u)\,\mathrm{P}_{l}^{-(u+1}(\mu)\,e^{i(u+1)\phi}
\end{aligned}\right\}\; -(l + 1) \leqslant u \leqslant -1 \qquad (19\text{B})$$

‡

where F_l, G_l are functions of r only and $\mu = \cos\theta$; the other, corresponding to $(l, j = l - \frac{1}{2}, m = u + \frac{1}{2})$ given by

$$\left.\begin{aligned}
\psi_1 &= -\,i\mathrm{F}_{-l-1}\,(l + u)\,\mathrm{P}_{l-1}^{u}(\mu)\,e^{iu\phi} \\
\psi_2 &= i\mathrm{F}_{-l-1}\,\mathrm{P}_{l-1}^{u+1}(\mu)\,e^{i(u+1)\phi} \\
\psi_3 &= \mathrm{G}_{-l-1}\,(l - u)\,\mathrm{P}_{l}^{u}(\mu)\,e^{iu\phi} \\
\psi_4 &= \mathrm{G}_{-l-1}\,\mathrm{P}_{l}^{u+1}(\mu)\,e^{i(u+1)\phi}
\end{aligned}\right\}\; 0 \leqslant u \leqslant l - 1 \qquad (20\text{A})$$

$$\left.\begin{aligned}
\psi_1 &= i\mathrm{F}_{-l-1}\,\mathrm{P}_{l-1}^{-u}(\mu)\,e^{iu\phi} \\
\psi_2 &= i\mathrm{F}_{-l-1}\,(l - u - 1)\,\mathrm{P}_{l-1}^{-(u+1)}(\mu)\,e^{i(u+1)\phi} \\
\psi_3 &= -\,\mathrm{G}_{-l-1}\,\mathrm{P}_{l}^{-u}(\mu)\,e^{iu\phi} \\
\psi_4 &= \mathrm{G}_{-l-1}\,(l + u + 1)\,\mathrm{P}_{l}^{-(u+1}(\mu)\,e^{i(u+1)\phi}
\end{aligned}\right\}\; -l \leqslant u \leqslant -1. \qquad (20\text{B})$$

The functions F_l, G_l are solutions of the differential equations

$$\left.\begin{aligned}
&\left(-\frac{\varepsilon + \mathrm{V}}{c} + c\right)\mathrm{F}_l + \frac{d\mathrm{G}_l}{dr} - \frac{l}{r}\mathrm{G}_l = 0 \\
&\left(\frac{\varepsilon + \mathrm{V}}{c} + c\right)\mathrm{G}_l + \frac{d\mathrm{F}_l}{dr} + \frac{l + 2}{r}\mathrm{F}_l = 0
\end{aligned}\right\}. \qquad (21)$$

Putting

$$\mathrm{P}_l = r\mathrm{G}_l \quad , \quad \mathrm{Q}_l = r\mathrm{F}_l \qquad (22)$$

---

† *Cf.* D. R. Hartree, 'Proc. Camb. Phil. Soc.,' vol. 25, p. 225, quoted as A.

‡ *See* footnote †, p. 629.

equations (21) become

$$\left.\begin{aligned}\left(-\frac{\varepsilon + \mathrm{V}}{c} + c\right)\mathrm{Q}_l + \frac{d\mathrm{P}_l}{dr} - \frac{l+1}{r}\mathrm{P}_l = 0\\ \left(\frac{\varepsilon + \mathrm{V}}{c} + c\right)\mathrm{P}_l + \frac{d\mathrm{Q}_l}{dr} + \frac{l+1}{r}\mathrm{Q}_l = 0\end{aligned}\right\} \quad . \tag{23}$$

The second-order equation derived for P_l from these equations is, when the relativistic terms are neglected, the same as the non-relativistic equation for the function P_l defined as r times the radial wave function, *i.e.*, P_l corresponds to the non-relativistic P_l used by Hartree.† We could at once build up the relativistic self-consistent field equation without "exchange" terms by taking for V the sum of the potential due to a nucleus of charge Z and that of the charge distribution $\Sigma\psi_\rho{}^*\psi_\rho$ where the summation is taken over the electrons in other wave functions than the particular one considered. In what follows we show how to derive the relativistic self-consistent field equations including "exchange" terms.

3—We follow the line suggested by Slater for the non-relativistic case for an atom with N electrons.

We first construct an expression for the total energy of the system in terms of one-electron wave functions. Let the expression

$$\psi_{\rho_s}(\alpha_p \mid s), \qquad (\rho_s = 1 \ldots 4)$$

denote the wave function for the state α_p, specified by the four quantum numbers (n, l, j, m), expressed as a function of the co-ordinates of the sth electron. That is, the sth electron is in the state α_p. We consider the integral

$$\int \sum_{\substack{\rho_1\ldots\rho_N=1\\ \sigma_1\ldots\sigma_N=1}}^{4} \sum_{\mathrm{P}}(-1)^{\mathrm{P}}\,\psi_{\rho_1}{}^*(\alpha_1 \mid 1)\ldots\psi_{\rho_N}{}^*(\alpha_N \mid \mathrm{N})\,\mathrm{H}_{\rho_1\ldots\rho_N,\,\sigma_1\ldots\sigma_N} \sum_{\mathrm{P}}(-1)^{\mathrm{P}}\,\psi_{\sigma_1}(\beta_1 \mid 1)\ldots\psi_{\sigma_N}(\beta_N \mid \mathrm{N})\,d\tau_1\ldots d\tau_{\mathrm{N}}, \tag{24}$$

where $\mathrm{H}_{\rho_1\ldots\rho_N,\,\sigma_1\ldots\sigma_N}$, the hamiltonian for the whole atom is the sum of several terms; firstly, the hamiltonians of the separate electrons in the field of the nucleus, for example, that of electron 1, namely,

$$\left(-\frac{\mathrm{Z}}{r_1}\mathrm{U}_{\rho_1\sigma_1} + c\boldsymbol{\alpha}_{\rho_1\sigma_1}\cdot\mathbf{p}_1 + c^2\beta_{\rho_1\sigma_1}\right)\mathrm{U}_{\rho_2\sigma_2}\ldots\mathrm{U}_{\rho_N\sigma_N}, \tag{25}$$

and secondly, terms due to the interaction between pairs of electrons, which, as we have pointed out, we take to be the crude electrostatic interaction, so that these terms are of the type

$$\frac{1}{r_{12}}\,\mathrm{U}_{\rho_1\sigma_1}\ldots\mathrm{U}_{\rho_N\sigma_N}, \tag{26}$$

† No confusion should arise between P_l and the Legendre functions.

where r_{12} is the distance between electrons 1 and 2. The summation $\sum\limits_{P}(-1)^{P}$ means that the summation is to be taken over all permutations of the electrons among the different wave functions, the positive or negative sign to be taken according as the permutation is even or odd.

We contemplate only wave functions constructed of ψ's of the forms (19) and (20) as regards their dependence on θ and ϕ. As in the non-relativistic case, we work with orthogonal one-electron wave functions. We determine the differential equations for the radial wave functions P, Q, by making (24) take a stationary value for variations of P and Q subject to the normalizing and orthogonalizing conditions for the one-electron wave functions. The normalizing factors for the solutions (19*a*), (19*b*), (20*a*), (20*b*) are respectively

$$\left[\frac{1}{4\pi}\frac{(l-u)\,!}{(l+u+1)\,!}\right]^{\frac{1}{2}}\left[\frac{1}{4\pi}\frac{(l+u+1)\,!}{(l-u)\,!}\right]^{\frac{1}{2}}\left[\frac{1}{4\pi}\frac{(l-u-1)\,!}{(l+u)\,!}\right]^{\frac{1}{2}}\left[\frac{1}{4\pi}\frac{(l+u)\,!}{(l-u-1)\,!}\right]^{\frac{1}{2}}, \tag{27}$$

with

$$\int_0^\infty (\mathrm{F}_l^2+\mathrm{G}_l^2)\,r^2\,dr = \int_0^\infty (\mathrm{P}_l^2+\mathrm{Q}_l^2)\,dr$$
$$= \int_0^\infty (\mathrm{F}_{-l-1}^2+\mathrm{G}_{-l-1}^2)\,r^2\,dr = \int_0^\infty (\mathrm{P}_{-l-1}^2+\mathrm{Q}_{-l-1}^2)\,dr = 1. \tag{28}$$

We therefore multiply the solutions (19) and (20) by the appropriate factors (27) and the normalizing condition for the radial wave functions $\mathrm{P}_{n,\,l}$, $\mathrm{Q}_{n,\,l}$ for a state $(n, l, j = l+\frac{1}{2}, m)$ is

$$\int_0^\infty (\mathrm{P}_{n,\,l}^2+\mathrm{Q}_{n,\,l}^2)\,dr = 1, \tag{29}$$

while that for $\mathrm{P}_{n,\,-l-1}$, $\mathrm{Q}_{n,\,-l-1}$, corresponding to a state $(n, l, j = l-\frac{1}{2}, m)$, is

$$\int_0^\infty (\mathrm{P}_{n,\,-l-1}^2+\mathrm{Q}_{n,\,-l-1}^2)\,dr = 1. \tag{30}$$

Two wave functions of the types (19) and (20) are orthogonal (that is to say, $\int \psi_{p_s}(\alpha_p \mid s)\,\psi_{p_s}(\alpha_q \mid s)\,d\tau = 0$), unless the triad of quantum numbers (l, j, m) is the same for both; for two functions corresponding to states (n, l, j, m), (n', l, j, m) the orthogonalizing condition is

$$\int (\mathrm{P}_{n,\,\substack{l\\-l-1}}\mathrm{P}_{n',\,\substack{l\\-l-1}}+\mathrm{Q}_{n,\,\substack{l\\-l-1}}\mathrm{Q}_{n',\,\substack{l\\-l-1}})\,dr = 0. \tag{31}$$

We proceed to the evaluation of (24) in terms of integrals involving these radial wave functions only. The terms of the type (25) involving the co-ordinates of a single electron only give, for completed sub-groups,

$$\begin{aligned}
\mathrm{N}!\sum_{n,\,l\geq 0,\,j=l+\frac{1}{2}}(2j+1)\int_0^\infty\Big[&-\Big(\frac{Z}{cr}+c\Big)Q_{n,l}^2+Q_{n,l}\Big(\frac{dP_{n,l}}{dr}-\frac{l+1}{r}P_{n,l}\Big)\\
&-\Big(\frac{Z}{cr}+c\Big)P_{n,l}^2-P_{n,l}\Big(\frac{dQ_{n,l}}{dr}+\frac{l+1}{r}Q_{n,l}\Big)\Big]dr
\end{aligned}$$

$$\begin{aligned}
&+\mathrm{N}!\sum_{n,\,l\geq 1,\,j=l-\frac{1}{2}}(2j+1)\times\\
&\quad\int_0^\infty\Big[-\Big(\frac{Z}{cr}+c\Big)Q_{n,-l-1}^2+Q_{n,-l-1}\Big(\frac{dP_{n,-l-1}}{dr}+\frac{l}{r}P_{n,-l-1}\Big)\\
&\qquad-\Big(\frac{Z}{cr}+c\Big)P_{n,-l-1}^2-P_{n,-l-1}\Big(\frac{dQ_{n,-l-1}}{dr}-\frac{l}{r}Q_{n,-l-1}\Big)\Big]dr. \qquad (32)
\end{aligned}$$

The contributions from terms of the type (26) will be of two kinds, the direct interaction integrals for which the permutation among the ψ^*'s is the same as that among the ψ's and the "exchange" integrals for which the permutation of the electrons among the ψ^*'s differs from that among the ψ's by the interchange of two electrons. As in the non-relativistic case, all other permutations give zero contributions when the orthogonality relations are satisfied. These terms therefore give altogether

$$\frac{\mathrm{N}!}{c}\sum_{\substack{n,l,j,u\\ n',l',j',u'}}[\mathrm{J}(n,l,j,u;n',l',j',u')-\mathrm{K}(n,l,j,u;n',l',j',u')], \qquad (33)$$

where J is the direct integral, K the exchange integral. Now,

$$\frac{1}{r_{12}}=\sum_k\sum_{m=-k}^{k}\frac{(k-|m|)!}{(k+|m|)!}\frac{r(a)^k}{r(b)^{k+1}}P_k^{|m|}(\mu_1)P_k^{|m|}(\mu_2)e^{im(\phi_1-\phi_2)}, \qquad (34)$$

where $r(a)$ is the smaller and $r(b)$ the greater of r_1 and r_2 and $\mu_1=\cos\theta_1$ $\mu_2=\cos\theta_2$. We have, therefore, for $u, u'\geqslant 0$

$$\begin{aligned}
&\mathrm{J}\Big(n,l,l+\tfrac{1}{2},\substack{u\\ -(u+1)};n',l',l'+\tfrac{1}{2},\substack{u'\\ -(u'+1)}\Big)\\
&=\frac{1}{4\pi^3}\frac{(l-u)!}{(l+u+1)!}\frac{(l'-u')}{(l'+u'+1)!}\\
&\quad\times\sum_k\iint\{Q_{n,l}^2(r_1)[(l-u+1)^2(P_{l+1}^u(\mu_1))^2+(P_{l+1}^{u+1}(\mu_1))^2]\\
&\qquad+P_{n,l}^2(r_1)[(l+u+1)^2(P_l^u(\mu_1))^2+(P_l^{u+1}(\mu_1))^2]\}\\
&\quad\times\{Q_{n',l'}^2(r_2)[(l'-u'+1)^2(P_{l'+1}^{u'}(\mu_2))^2+(P_{l'+1}^{u'+1}(\mu_2))^2]\\
&\qquad+P_{n',l'}^2(r_2)[(l'+u'+1)^2(P_{l'}^{u'}(\mu_2))^2+(P_{l'}^{u'+1}(\mu_2))^2]\}\\
&\qquad\times P_k^0(\mu_1)P_k^0(\mu_2)\frac{r(a)^k}{r(b)^{k+1}}d\tau_1\,d\tau_2 \qquad (35)\\
&=\sum a_k\Big(l,l+\tfrac{1}{2},\substack{u\\ -(u+1)};l',l'+\tfrac{1}{2},\substack{u'\\ -(u'+1)}\Big)\mathrm{F}_k(n,l;n',l'), \qquad (36)
\end{aligned}$$

where

$$a_k\left(l, l+\tfrac{1}{2}, \overset{u}{-(u+1)}; l', l'+\tfrac{1}{2}, \overset{u'}{-(u'+1)}\right)$$

$$= \frac{(l-u)!\,(l'-u')!}{(l+u+1)!\,(l'+u'+1)!} \int_{-1}^{1} [(l+u+1)^2 (\mathrm{P}_l^u(\mu_1))^2 + (\mathrm{P}_l^{u+1}(\mu_1))^2]\, \mathrm{P}_k^0(\mu_1) \frac{d\mu_1}{2}$$
$$\times \int_{-1}^{1} [(l'+u'+1)^2 (\mathrm{P}_{l'}^{u'}(\mu_2))^2 + (\mathrm{P}_{l'}^{u'+1}(\mu_2))^2]\, \mathrm{P}_k^0(\mu_2) \frac{d\mu_2}{2} \tag{37}$$

and

$$\mathrm{F}_k(n, l; n', l') = \iint [\mathrm{P}_{n,l}^2(r_1)\,\mathrm{P}_{n',l'}^2(r_2) + \mathrm{P}_{n,l}^2(r_1)\,\mathrm{Q}_{n',l'}^2(r_2) + \mathrm{Q}_{n,l}^2(r_1)\,\mathrm{P}_{n',l'}^2(r_2) + \mathrm{Q}_{n,l}^2(r_1)\,\mathrm{Q}_{n',l'}^2(r_2)] \times \frac{r(a)^k}{r(b)^{k+1}}\, dr_1\, dr_2. \tag{38}$$

The proof that in (35) the coefficients of all the different products of a given set of $\mathrm{P}_{n,l}$, $\mathrm{Q}_{n,l}$, $\mathrm{P}_{n',l'}$, $\mathrm{Q}_{n',l'}$ are equal is given in Hartree's paper A. The notation is analogous to that used by Slater.

We find similarly that

$$\mathrm{J}\left(n, l, l-\tfrac{1}{2}, \overset{u}{-(u+1)}; n', l', l'-\tfrac{1}{2}, \overset{u'}{-(u'+1)}\right)$$
$$= \Sigma\, a_k\left(l, l-\tfrac{1}{2}, \overset{u}{-(u+1)}; l', l'-\tfrac{1}{2}, \overset{u'}{-(u'+1)}\right) \mathrm{F}_k(n, -(l+1);\ n', -(l'+1)),$$

where

$$a_k\left(l, l-\tfrac{1}{2}, \overset{u}{-(u+1)}; l', l'-\tfrac{1}{2}, \overset{u'}{-(u'+1)}\right)$$
$$= a_k\left(l-1, l-\tfrac{1}{2}, \overset{u}{-(u+1)}; l'-1, l'-\tfrac{1}{2}, \overset{u'}{-(u'+1)}\right) \tag{39}$$

and

$$\mathrm{J}\left(n, l, l+\tfrac{1}{2}, \overset{u}{-(u+1)}; n', l', l'-\tfrac{1}{2}, \overset{u'}{-(u'+1)}\right)$$
$$= \Sigma a_k\left(l, l+\tfrac{1}{2}, \overset{u}{-(u+1)}; l', l'-\tfrac{1}{2}, \overset{u'}{-(u'+1)}\right) \mathrm{F}_k(n, l; n', -(l'+1)), \tag{40}$$

where

$$a_k\left(l, l+\tfrac{1}{2}, \substack{u \\ -(u+1)}; l', l'-\tfrac{1}{2}, \substack{u' \\ -(u'+1)}\right)$$
$$= a_k\left(l, l+\tfrac{1}{2}, \substack{u \\ -(u+1)}; l'-1, l'-\tfrac{1}{2}, \substack{u' \\ -(u'+1)}\right). \tag{41}$$

In all these formulae the four values, u, $-(u+1)$; u', $-(u'+1)$ can be taken together in any of the four possible ways. The table of a's can be very simply constructed from the set given by (37). Table I gives the values of the a's for all pairs of s, p, and d wave functions.

We now proceed to the evaluation of the "exchange" integrals K. If $u, u' \geqslant 0$

$$\mathrm{K}\left(n, l, l+\tfrac{1}{2}, \substack{u \\ -(u+1)}; n', l', l'+\tfrac{1}{2}, \substack{u' \\ -(u'+1)}\right)$$
$$= \Sigma\, b_k\left(l, l+\tfrac{1}{2}, \substack{u \\ -(u+1)}; l', l'+\tfrac{1}{2}, \substack{u' \\ -(u+1)}\right)\mathrm{G}_k(n, l; n', l'), \tag{42}$$

where

$$b_k\left(l, l+\tfrac{1}{2}, \substack{u \\ -(u+1)}; l', l'+\tfrac{1}{2} \substack{u' \\ -(u'+1)}\right)$$
$$= \frac{(l-u)\,!}{(l+u+1)} \frac{(l'-u')\,!}{(l'+u'+1)} \frac{(k-|u'-u|)\,!}{(k+|u'-u|)\,!}$$
$$\times \left[\int_{-1}^{1} \{(l+u+1)(l'+u'+1)\,\mathrm{P}_l^u(\mu)\,\mathrm{P}_{l'}^{u'}(\mu)\right.$$
$$\left. + \mathrm{P}_l^{u+1}(\mu)\,\mathrm{P}_{l'}^{u'+1}(\mu)\}\,\mathrm{P}_k^{|u'-u|}(\mu)\,\frac{d\mu}{2}\right]^2 \tag{43}$$

and

$$\mathrm{G}_k(n, l; n', l')$$
$$= \iint [\mathrm{P}_{n,l}(r_1)\,\mathrm{P}_{n',l'}(r_1)\,\mathrm{P}_{n,l}(r_2)\,\mathrm{P}_{n',l'}(r_2) + \mathrm{P}_{n,l}(r_1)\,\mathrm{P}_{n',l'}(r_1)\,\mathrm{Q}_{n,l}(r_2)\,\mathrm{Q}_{n',l'}(r_2)$$
$$+ \mathrm{Q}_{n,l}(r_1)\mathrm{Q}_{n',l'}(r_1)\mathrm{P}_{n,l}(r_1)\mathrm{P}_{n',l'}(r_1) + \mathrm{Q}_{n,l}(r_1)\,\mathrm{Q}_{n',l'}(r_1)\,\mathrm{Q}_{n,l}(r_2)\,\mathrm{Q}_{n',l'}(r_2)]$$
$$\times \frac{r(a)^k}{r(b)^{k+1}}\,dr_1\,dr_2. \tag{44}$$

Here u is to be taken with u' and $-(u+1)$ with $-(u'+1)$.

The proof that the coefficients of all the different products of a given set of $\mathrm{P}_{n,l}$, $\mathrm{P}_{n',l'}$ $\mathrm{Q}_{n,l}$, $\mathrm{Q}_{n',l'}$ are the same follows very simply from the properties of spherical harmonics.†

† *See* Hartree, A, p. 228 (7) and (8).

Further

$$\mathrm{K}\left(n, l, l+\tfrac{1}{2}, \substack{u \\ -(u+1)}; n', l', l'+\tfrac{1}{2}, \substack{-(u'+1) \\ u'}\right)$$

$$= \Sigma_k \left(b_k\, l, l+\tfrac{1}{2}, \substack{u \\ -(u+1)};\ l', l'+\tfrac{1}{2}, \substack{-(u'+1) \\ u'}\right) \mathrm{G}_k(n, l; n', l'), \quad (45)$$

where

$$b_k\left(l, l+\tfrac{1}{2}, \substack{u \\ -(u+1)}; l', l'+\tfrac{1}{2}, \substack{-(u'+1) \\ u'}\right)$$

$$= \frac{(l-u)\,!}{(l+u+1)\,!} \frac{(l'-u')\,!}{(l'+u'+1)\,!} \frac{(k-|u'+u+1|)}{(k+|u'+u+1|)}$$

$$\times \left[\int_{-1}^{1} \{(l'+u'+1)\, \mathrm{P}_l^{u+1}(\mu)\, \mathrm{P}_{l'}^{u'}(\mu) - (l+u+1)\, \mathrm{P}_l^{u}(\mu)\, \mathrm{P}_{l'}^{u'+1}(\mu)\}\right.$$

$$\left. \times\ \mathrm{P}_k^{|u'+u+1|}(\mu) \frac{d\mu}{2}\right]^2. \quad (46)$$

Again the two upper and two lower signs are to be taken together.

We find also that

$$\mathrm{K}\left(n, l, l-\tfrac{1}{2}, \substack{u \\ -(u+1)}; n', l', l'-\tfrac{1}{2}, \substack{u' \\ -(u'+1)}\right)$$

$$= \Sigma b_k\left(l, l-\tfrac{1}{2}, \substack{u \\ -(u+1)}; l', l'-\tfrac{1}{2}, \substack{u' \\ -(u'+1)}\right)$$

$$\mathrm{G}_k(n, -(l+1); n', -(l'+1)), \quad (47)$$

where

$$b_k\left(l, l-\tfrac{1}{2}, \substack{u \\ -(u+1)}; l', l'-\tfrac{1}{2}, \substack{u' \\ -(u'+1)}\right)$$

$$= b_k\left(l-1, l-\tfrac{1}{2}, \substack{u \\ -(u+1)}; l'-1, l'-\tfrac{1}{2}, \substack{u' \\ -(u'+1)}\right) \quad (48)$$

for all four possible combinations of u, $-(u+1)$; u', and $-(u'+1)$. There are two remaining different types of integral,

$$\mathrm{K}\left(n, l, l+\tfrac{1}{2}, \substack{u \\ -(u+1)}; n', l', l'-\tfrac{1}{2}, \substack{u' \\ -(u'+1)}\right)$$

and

$$\mathrm{K}\left(n, l, l+\tfrac{1}{2}, \substack{u \\ -(u+1)}; n', l', l'-\tfrac{1}{2}, \substack{-(u'+1) \\ u'}\right)$$

where, in each case, the two upper and two lower signs are to be taken together. For the first,

$$\mathrm{K}\left(n, l, l+\tfrac{1}{2}, \substack{u \\ -(u+1)}; n', l', l'-\tfrac{1}{2}, \substack{u' \\ -(u'+1)}\right)$$

$$= \Sigma b_k\left(l, l+\tfrac{1}{2}, \substack{u \\ -(u+1)}; l', l'-\tfrac{1}{2}, \substack{u' \\ -(u'+1)}\right)$$

$$\mathrm{G}_k(n, l; n', -(l'+1)), \quad (49)$$

where

$$b_k\left(l,\, l+\tfrac{1}{2},\, \begin{smallmatrix} u \\ -(u+1) \end{smallmatrix};\ l',\, l'-\tfrac{1}{2},\, \begin{smallmatrix} u' \\ -(u'+1) \end{smallmatrix}\right)$$
$$= \frac{(l-u)!}{(l+u+1)!}\,\frac{(l'-u'-1)!}{(l'+u')!}\,\frac{(k-|u'-u|)!}{(k+|u'-u|)!}$$
$$\times \left[\int_{-1}^{1} \{(l+u+1)(l'-u')\,\mathrm{P}_l^u(\mu)\,\mathrm{P}_{l'}^{u'}(\mu) - \mathrm{P}_l^{u+1}(\mu)\,\mathrm{P}_{l'}^{u'+1}(\mu)\}\,\mathrm{P}_k^{|u'-u|}(\mu)\,\frac{d\mu}{2}\right]^2. \tag{50}$$

For the second

$$\mathrm{K}\left(n,\, l,\, l+\tfrac{1}{2},\, \begin{smallmatrix} u \\ -(u+1) \end{smallmatrix};\ n',\, l',\, l'-\tfrac{1}{2},\, \begin{smallmatrix} -(u'+1) \\ u' \end{smallmatrix}\right)$$
$$= \Sigma b_k\left(l,\, l+\tfrac{1}{2},\, \begin{smallmatrix} u \\ -(u+1) \end{smallmatrix};\ l',\, l'-\tfrac{1}{2},\, \begin{smallmatrix} -(u'+1) \\ u' \end{smallmatrix}\right)\mathrm{G}\,(n, l;\, n', l'), \tag{51}$$

where

$$b_k\left(l,\, l+\tfrac{1}{2},\, \begin{smallmatrix} u \\ -(u+1) \end{smallmatrix};\ l',\, l'-\tfrac{1}{2},\, \begin{smallmatrix} -(u'+1) \\ u' \end{smallmatrix}\right)$$
$$= \frac{(l-u)!}{(l+u+1)!}\,\frac{(l'-u'-1)!}{(l'+u')!}\,\frac{(k-|u'+u+1|)!}{(k+|u'+u+1|)!}$$
$$\left[\int_{-1}^{1} \{(l'-u')\,\mathrm{P}_l^{u+1}(\mu)\,\mathrm{P}_{l'}^{u'}(\mu) + (l+u+1)\,\mathrm{P}_l^{u}(\mu)\,\mathrm{P}_{l'}^{u'+1}(\mu)\}\,\mathrm{P}_k^{|u'+u+1|}(\mu)\,\frac{d\mu}{2}\right]^2. \tag{52}$$

The four different types of *b* given by (43), (46), (50), and (52) have been evaluated for all pairs of *s*, *p*, *d* wave functions and the values of all the *b*'s derivable from them are given in Table II.

To show the relation of our results to the corresponding results of the non-relativistic treatment, it is desirable to adapt the conventional notation which uses the letters *s*, *p*, *d*, ... for wave functions with $l = 0, 1, 2, \ldots$; we shall distinguish those for which $j = l - \frac{1}{2}$ by a bar, thus:

	s	*p*	$\bar{p}$	*d*	$\bar{d}$
l	0	1	1	2	2
j	1/2	3/2	1/2	5/2	3/2

In Table I the groups of values for *u*, *u'* may be taken together in any of the possible combinations. In Table II only the values on the same level are to be taken together.

4—We now construct a table of the total coefficients of the $\mathrm{F}_k\,(n, l;\, n', l')$ and $\mathrm{G}_k\,(n, l;\, n', l')$ in the expression (33) for the interaction energy, for

TABLE I—VALUES OF a_k

Electrons	l	j	u	l'	j'	u'	a_k $k = 0$	$k = 2$	$k = 4$
ss	0	1/2	0 −1	0	1/2	0 −1	1	—	—
sp	0	1/2	0 −1	1	3/2	0, 1 −1, −2	1	—	—
$s\bar{p}$	0	1/2	0 −1	1	1/2	0 −1	1	—	—
sd	0	1/2	0 −1	2	5/2	0, 1, 2 −1, −2, −3	1	—	—
$s\bar{d}$	0	1/2	0 −1	2	3/2	0, 1 −1, −2	1	—	—
pp	1	3/2	0 −1	1	3/2	0 −1	1	1/25	—
	1	3/2	0 −1	1	3/2	1 −2	1	−1/25	—
	1	3/2	1 −2	1	3/2	1 −2	1	1/25	—
$p\bar{p}, \bar{p}\bar{p}$	1	1/2	0 −1	1	3/2, 1/2	0, 1 −1, −2	1	—	—
pd	1	3/2	0 −1	2	5/2	0 −1	1	8/175	—
	1	3/2	0 −1	2	5/2	1 −2	1	2/175	—
	1	3/2	0 −1	2	5/2	2 −3	1	−10/175	—
	1	3/2	1 −2	2	5/2	0 −1	1	−8/175	—
	1	3/2	1 −2	2	5/2	1 −2	1	−2/175	—
	1	3/2	1 −2	2	5/2	2 −3	1	10/175	—

$p\bar{d}$	1	3/2	0 −1	2	3/2	0 −1	1	1/25	—
	1	3/2	0 −1	2	3/2	1 −2	1	−1/25	—
	1	3/2	1 −2	2	3/2	0 −1	1	−1/25	—
	1	3/2	1 −2	2	3/2	1 −2	1	1/25	—
$\bar{p}d$	1	1/2	0 −1	2	5/2	0, 1, 2 −1, −2, −3	1	—	—
$\bar{p}\bar{d}$	1	1/2	0 −1	2	3/2	0, 1 −1, −2	1	—	—
dd	2	5/2	0 −1	2	5/2	0 −1	1	64/1225	4/441
	2	5/2	0 −1	2	5/2	1 −2	1	16/1225	−6/441
	2	5/2	0 −1	2	5/2	2 −3	1	−80/1225	2/441
	2	5/2	1 −2	2	5/2	1 −2	1	4/1225	9/441
	2	5/2	1 −2	2	5/2	2 −3	1	−20/1225	−3/441
	2	5/2	2 −3	2	5/2	2 −3	1	100/1225	1/441
$\bar{d}\bar{d}$	2	3/2	0 −1	2	3/2	0 −1	1	1/25	—
	2	3/2	0 −1	2	3/2	1 −2	1	−1/25	—
	2	3/2	1 −2	2	3/2	1 −2	1	1/25	—
$d\bar{d}$	2	3/2	0 −1	2	5/2	0 −1	1	8/175	—
	2	3/2	0 −1	2	5/2	1 −2	1	2/175	—
	2	3/2	0 −1	2	5/2	2 −3	1	−10/175	—
	2	3/2	1 −2	2	5/2	0 −1	1	−8/175	—
	2	3/2	1 −2	2	5/2	1 −2	1	−2/175	—
	2	3/2	1 −2	2	5/2	2 −3	1	10/175	—

TABLE II—VALUES OF b_k

Electrons	l	j	u	l'	j'	u'	b_k				
							$k=0$	1	2	3	4
ss	0	1/2	0 −1	0	1/2	0 −1	1	—	—	—	—
	0	1/2	0	0	1/2	−1	—	—	—	—	—
sp	0	1/2	0 −1	1	3/2	0 −1	—	2/9	—	—	—
	0	1/2	0 −1	1	3/2	1 −2	—	3/9	—	—	—
	0	1/2	0 −1	1	3/2	−1 0	—	1/9	—	—	—
	0	1/2	0 −1	1	3/2	−2 1	—	—	—	—	—
sp	0	1/2	0 −1	1	1/2	0 −1	—	1/9	—	—	—
	0	1/2	0 −1	1	1/2	1 −2	—	—	—	—	—
	0	1/2	0 −1	1	1/2	−1 0	—	2/9	—	—	—
	0	1/2	0 −1	1	1/2	−2 1	—	—	—	—	—
sd	0	1/2	0 −1	2	5/2	0 −1	—	—	3/25	—	—
	0	1/2	0 −1	2	5/2	1 −2	—	—	4/25	—	—
	0	1/2	0 −1	2	5/2	2 −3	—	—	5/25	—	—
	0	1/2	0 −1	2	5/2	−1 0	—	—	2/25	—	—
	0	1/2	0 −1	2	5/2	−2 1	—	—	1/25	—	—

sd	0	1/2	0 −1	2	5/2	−3 2	—	—	—	—	—
sd̄	0	1/2	0 −1	2	3/2	0 −1	—	—	1/25	—	—
	0	1/2	0 −1	2	3/2	1 −2	—	—	2/25	—	—
	0	1/2	0 −1	2	3/2	−1 0	—	—	3/25	—	—
	0	1/2	0 −1	2	3/2	−2 1	—	—	4/25	—	—
pp	1	3/2	0 −1	1	3/2	0 −1	1	—	1/25	—	—
	1	3/2	0 −1	1	3/2	1 −2	—	—	2/25	—	—
	1	3/2	1 −2	1	3/2	1 −2	1	—	1/25	—	—
	1	3/2	0 −1	1	3/2	−1 0	—	—	—	—	—
	1	3/2	0 −1	1	3/2	−2 1	—	—	2/25	—	—
	1	3/2	1 −2	1	3/2	−2 1	—	—	—	—	—
p̄p̄	1	1/2	0 −1	1	1/2	0 −1	1	—	—	—	—
	1	1/2	0 −1	1	1/2	−1 0	—	—	—	—	—
pp̄	1	3/2	0 −1	1	1/2	0 −1	—	—	2/25	—	—
	1	3/2	1 −2	1	1/2	0 −1	—	—	1/25	—	—
	1	3/2	0 −1	1	1/2	−1 0	—	—	3/25	—	—
	1	3/2	1 −2	1	1/2	−1 0	—	—	4/25	—	—

Table II—(continued)

Electrons	l	j	u	l'	j'	u'	b_k				
							$k = 0$	1	2	3	4
pd	1	3/2	0 −1	2	5/2	0 −1	—	6/25	—	24/1225	—
	1	3/2	0 −1	2	5/2	1 −2	—	6/25	—	49/1225	—
	1	3/2	0 −1	2	5/2	2 −3	—	—	—	50/1225	—
	1	3/2	1 −2	2	5/2	0 −1	—	1/25	—	54/1225	—
	1	3/2	1 −2	2	5/2	1 −2	—	4/25	—	36/1225	—
	1	3/2	1 −2	2	5/2	2 −3	—	10/25	—	15/1225	—
	1	3/2	0 −1	2	5/2	−1 0	—	3/25	—	2/1225	—
	1	3/2	0 −1	2	5/2	−2 1	—	—	—	10/1225	—
	1	3/2	0 −1	2	5/2	−3 2	—	—	—	75/1225	—
	1	3/2	1 −2	2	5/2	−1 0	—	—	—	60/1225	—
	1	3/2	1 −2	2	5/2	−2 1	—	—	—	45/1225	—
	1	3/2	1 −2	2	5/2	−3 2	—	—	—	—	—
$\overline{p\,d}$	1	1/2	0 −1	2	3/2	0 −1	—	2/9	—	—	—
	1	1/2	0 −1	2	3/2	1 −2	—	3/9	—	—	—
	1	1/2	0 −1	2	3/2	−1 0	—	1/9	—	—	—
	1	1/2	0 −1	2	3/2	−2 1	—	—	—	—	—

$p\bar{d}$	1	3/2	0 −1	2	3/2	0 −1	—	1/225	—	81/1225	—
	1	3/2	0 −1	2	3/2	1 −2	—	6/225	—	36/1225	—
	1	3/2	1 −2	2	3/2	0 −1	—	6/225	—	36/1225	—
	1	3/2	1 −2	2	3/2	1 −2	—	9/225	—	9/1225	—
	1	3/2	0 −1	2	3/2	−1 0	—	8/225	—	108/1225	—
	1	3/2	0 −1	2	3/2	−2 1	—	—	—	90/1225	—
	1	3/2	1 −2	2	3/2	−1 0	—	—	—	90/1225	—
	1	3/2	1 −2	2	3/2	−2 1	—	—	—	180/1225	—
$\bar{p}d$	1	1/2	0 −1	2	5/2	0 −1	—	—	—	75/1225	—
	1	1/2	0 −1	2	5/2	1 −2	—	—	—	50/1225	—
	1	1/2	0 −1	2	5/2	2 −3	—	—	—	25/1225	—
	1	1/2	0 −1	2	5/2	−1 0	—	—	—	100/1225	—
	1	1/2	0 −1	2	5/2	−2 1	—	—	—	125/1225	—
	1	1/2	0 −1	2	5/2	−3 2	—	—	—	150/1225	—
dd	2	5/2	0 −1	2	5/2	0 −1	1	—	64/1225	—	4/441
	2	5/2	0 −1	2	5/2	1 −2	—	—	48/1225	—	10/441
	2	5/2	0 −1	2	5/2	2 −3	—	—	60/1225	—	9/441
	2	5/2	1 −2	2	5/2	1 −2	1	—	4/1225	—	9/441

2 x 2

Table II—(continued)

Electrons	l	j	u	l'	j'	u'	b_k				
							$k = 0$	1	2	3	4
dd	2	5/2	1 −2	2	5/2	2 −3	—	—	120/1225	—	4/441
	2	5/2	2 −3	2	5/2	2 −3	1	—	100/1225	—	1/441
	2	5/2	0 −1	2	5/2	−1 0	—	—	—	—	—
	2	5/2	0 −1	2	5/2	−2 1	—	—	108/1225	—	5/441
	2	5/2	0 −1	2	5/2	−3 2	—	—	—	—	14/441
	2	5/2	1 −2	2	5/2	−2 1	—	—	—	—	—
	2	5/2	1 −2	2	5/2	−3 2	—	—	—	—	14/441
	2	5/2	2 −3	2	5/2	−3 2	—	—	—	—	—
$\bar{d}\bar{d}$	2	3/2	0 −1	2	3/2	0 −1	1	—	1/25	—	—
	2	3/2	0 −1	2	3/2	1 −2	—	—	2/25	—	—
	2	3/2	1 −2	2	3/2	1 −2	1	—	1/25	—	—
	2	3/2	0 −1	2	3/2	−1 0	—	—	—	—	—
	2	3/2	0 −1	2	3/2	−2 1	—	—	2/25	—	—
	2	3/2	1 −2	2	3/2	−2 1	—	—	—	—	—

$d\bar{d}$	2	5/2	0 -1	2	3/2	0 -1	—	—	6/1225	—	24/441
	2	5/2	0 -1	2	3/2	1 -2	—	—	27/1225	—	10/441
	2	5/2	1 -2	2	3/2	0 -1	—	—	2/1225	—	15/441
	2	5/2	1 -2	2	3/2	1 -2	—	—	36/1225	—	4/441
	2	5/2	2 -3	2	3/2	0 -1	—	—	40/1225	—	6/441
	2	5/2	2 -3	2	3/2	1 -2	—	—	30/1225	—	1/441
	2	5/2	0 -1	2	3/2	-1 0	—	—	25/1225	—	30/441
	2	5/2	0 -1	2	3/2	-2 1	—	—	12/1225	—	20/441
	2	5/2	1 -2	2	3/2	-1 0	—	—	32/1225	—	30/441
	2	5/2	1 -2	2	3/2	-2 1	—	—	—	—	35/441
	2	5/2	2 -3	2	3/2	-1 0	—	—	—	—	21/441
	2	5/2	2 -3	2	3/2	-2 1	—	—	—	—	56/441

an atom built up of completed sub-groups, up to and including d and $\bar{d}$ electrons. With the aid of this it is possible to construct very simply the expression for the total energy of any atom not containing f or higher electrons. We denote a sub-group by (ns), (np) $(n\bar{p})$, etc. The numbers opposite (ns), for example, denote the coefficients of the $\mathrm{F}_k(n, 0; n, 0)$ or $\mathrm{G}_k(n, 0; n, 0)$ due to interactions within the (ns) shell. The numbers opposite (ns) $(n'p)$ denote the coefficients of the F_k $(n, 0; n', 1)$ or $\mathrm{G}_k(n, 0; n', 1)$ due to interactions between the (ns) and $(n'p)$ shells. In such expressions as (np) $(n'p)$, $n \neq n'$, but in those like (ns) $(n'p)$, n and n' can take all values.

TABLE III—COEFFICIENTS OF $\mathrm{F}_k(n, l; n', l')$ AND $\mathrm{G}_k(n, l; n', l')$

Sub-groups	Coefficient of $\mathrm{F}_k(n, l; n', l')$			Coefficient of $\mathrm{G}_k(n, l; n', l')$				
	$k = 0$	2	4	0	1	2	3	4
(ns)	1	—	—	—	—	—	—	—
(ns) $(n's)$	4	—	—	2	—	—	—	—
(ns) $(n'p)$	8	—	—	—	4/3	—	—	—
(ns) $(n'\bar{p})$	4	—	—	—	2/3	—	—	—
(ns) $(n'd)$	12	—	—	—	—	6/5	—	—
(ns) $(n'\bar{d})$	8	—	—	—	—	4/5	—	—
(np)	6	−2/25	—	—	—	8/25	—	—
$(n\bar{p})$	1	—	—	—	—	—	—	—
(np) $(n'\bar{p})$	8	—	—	—	—	4/5	—	—
(np) $(n'p)$	16	—	—	4	—	4/5	—	—
$(n\bar{p})$ $(n'\bar{p})$	4	—	—	2	—	—	—	—
(np) $(n'd)$	24	—	—	—	12/5	—	24/35	—
$(n\bar{p})$ $(n'd)$	12	—	—	—	—	—	6/7	—
(np) $(n'\bar{d})$	16	—	—	—	4/15	—	36/35	—
$(n\bar{p})$ $(n'\bar{d})$	8	—	—	—	4/3	—	—	—
(nd)	15	−24/175	−2/63	—	—	96/175	—	16/63
$(n\bar{d})$	6	−2/25	—	—	—	8/25	—	—
(nd) $(n'\bar{d})$	24	—	—	—	—	12/35	—	8/7
(nd) $(n'd)$	36	—	—	6	—	48/35	—	4/7
$(n\bar{d})$ $(n'\bar{d})$	16	—	—	4	—	4/5	—	—

In the non-relativistic case there is no distinction between the four quantities

$$\mathrm{F}_k\left(n, \begin{matrix} l \\ -(l+1) \end{matrix}; n', \begin{matrix} l' \\ -(l'+1) \end{matrix}\right) \quad \text{or} \quad \mathrm{G}_k\left(n, \begin{matrix} l \\ -(l+1) \end{matrix}; n', \begin{matrix} l' \\ -(l'+1) \end{matrix}\right).$$

If we put these four F's equal to each other, and similarly the four G's equal, the sums of the coefficients for the interaction of two completed *whole* groups are the same as the corresponding sums taken from Slater's tables. This provides a useful check on the calculations. The coefficients for interactions within a group are quite different from the non-relativistic ones, as we should expect, since the classification of wave functions within a group is quite different.

5—We now give an example of the derivation of the wave equations. The complete expression for the energy is the sum of terms of the types in (32) and (34). We obtain the equations for the radial wave functions for the $2p$ and $2\bar{p}$ states of neutral Cu (29). This is sufficient to bring out all the important features of the method.

In the expression for the total energy we vary $P_{2,1}$, $P_{2,-2}$, $Q_{2,1}$, $Q_{2,-2}$ separately, subject to the normalizing conditions (29) and (30) for $n = 2$, $l = 1$, and the orthogonal conditions (31) for $n = 2$, $l = 1$, $n = 3$, $l = 1$. Before performing the variation all integrals must be written in a form symmetrical in the two variables r_1 and r_2. The G's are already in such a form, but, when (n, l) (n', l') are different, $F_k(n, l; n', l')$ must be written as half the sum of (38) and (38) with r_1, r_2 interchanged. Similarly (32) must be re-written in a form symmetrical in r_1, r_2.

When $Q_{n,l}(r_1)$ only is varied,

$$\delta F_k(n, l;\ n', l')$$

$$= 2\iint Q_{n,l}(r_1)\,\delta Q_{n,l}(r_1)\,[P^2_{n',l'}(r_2) + Q^2_{n',l'}(r_2)]\,\frac{r(a)^k}{r(b)^{k+1}}\,dr_1\,dr_2. \quad (53)$$

We write

$$\int\frac{[P^2_{n',l'}(r_2) + Q^2_{n',l'}(r_2)]\,r(a)^k}{r(b)^{k+1}}\,dr_2 \quad \text{as} \quad V_k(n', l' \mid r_1). \quad (54)$$

Similarly

$$\delta G_k(n,l;\ n', l',) = \iint Q_{n',l'}(r_1)\,\delta Q_{n,l}(r_1)\,[P_{n,l}(r_2)\,P_{n',l'}(r_2)$$

$$+ Q_{n,l}(r_2)\,Q_{n',l'}(r_2)]\,\frac{r(a)^k}{r(b)^{k+1}}\,dr,\,dr_2, \quad (55)$$

and we write

$$\int[P_{n,l}(r_2)\,P_{n',l'}(r_2) + Q_{n,l}(r_2) + Q_{n',l'}(r_2)]\,\frac{r(a)^k}{r(b)^{k+1}} \quad \text{as} \quad V_k(n, l;\ n'\ l' \mid r_1). \quad (56)$$

We obtain similar expressions when $P_{n,l}(r_1)$ only is varied. The equations for the $2p$ wave-functions obtained by varying $Q_{2,1}(r_1)$ and $Q_{2,-2}(r_1)$ independently become

$$
\begin{aligned}
&\frac{dP_{2,1}}{dr} - \boxed{\frac{2}{r} P_{2,1}} + \left(c - \frac{\varepsilon_{(2,1)}}{c} - \frac{Z}{cr}\right) Q_{2,1} - \frac{\varepsilon_{(2,1;3,1)}}{c} Q_{3,1} \\
&+ \frac{1}{c}\Big[2 \sum_{n=1}^{3} V_0(n, 0 \mid r) + V_0(4, 0 \mid r) + 3V_0(2, 1 \mid r) + 2V_0(2, -2 \mid r) \\
&\qquad - \boxed{1/25\, V_2(2, 1 \mid r)} + 4V_0(3, 1 \mid r) + 2V_0(3, -2 \mid r) \\
&\qquad\qquad + 6V_0(3, 2 \mid r) + 4V_0(3, -3 \mid r)\Big] Q_{2,1} \\
&- \frac{1}{c}\Big[1/3 \sum_{n=1}^{3} V_1(n, 0; 2, 1 \mid r)\, Q(n, 0) + 1/6\, V_1(4, 0; 2, 1 \mid r)\, Q_{4,0} \\
&\quad + \boxed{2/25\, V_2(2, 1; 2, 1 \mid r)\, Q_{2,1} + 1/5\, V_2(2, 1; 2, -2 \mid r)\, Q_{2,-2}} \\
&\quad + \{V_0(3, 1; 2, 1 \mid r) + 1/5\, V_2(3, 1; 2, 1 \mid r)\}\, Q_{3,1} \\
&\quad + 1/5\, V_2(3, -2; 2, 1 \mid r)\, Q_{3,-2} \\
&\quad + \{3/5\, V_1(3, 2; 2, 1 \mid r) + 6/35\, V_3(3, 2; 2, 1 \mid r)\}\, Q_{3,2} \\
&\quad + \{1/15\, V_1(3, -3; 2, 1 \mid r) + 9/35\, V_3(3, -3; 2, 1 \mid r)\}\, Q_{3,-3}\Big] = 0 \quad (57)
\end{aligned}
$$

$$
\begin{aligned}
&\frac{dP_{2,-2}}{dr} + \boxed{\frac{1}{r} P_{2,-2}} + \left(c - \frac{\varepsilon_{(2,-2)}}{c} - \frac{Z}{cr}\right) Q_{2,-2} - \frac{\varepsilon_{(2,-2;3,-2)}}{c} Q_{3,-2} \\
&+ \frac{1}{c}\Big[2 \sum_{n=1}^{3} V_0(n, 0 \mid r) + V_0(4, 0 \mid r) + 4V_0(2, 1 \mid r) + V_0(2, -2 \mid r) \\
&\qquad + 4V_0(3, 1 \mid r) + 2V_0(3, -2 \mid r) + 6V_0(3, 2 \mid r) + 4V_0(3, -3 \mid r)\Big] Q_{2,-2} \\
&- \frac{1}{c}\Big[1/3 \sum_{n=1}^{3} V_1(n, 0; 2, 1 \mid r)\, Q_{n,0} + 1/6\, V_1(4, 0; 2, 1 \mid r)\, Q_{4,0} \\
&\quad + \boxed{2/5\, V_2(2, 1; 2, -2 \mid r)\, Q_{2,1}} + V_0(3, -2; 2, -2)\, Q_{3,-2} \\
&\quad + 2/5\, V_2(3, 1; 2, -2)\, Q_{3,1} + 2/3\, V_1(3, -3; 2, -2)\, Q_{3,-3} \\
&\qquad\qquad + 3/7\, V_3(3, 2; 2, -2)\, Q_{3,2}\Big] = 0 \quad (58)
\end{aligned}
$$

$\varepsilon_{(2,1)}$, $\varepsilon_{(2,-2)}$, $\varepsilon_{(2,1;3,1)}$ and $\varepsilon_{(2,-2;3,-2)}$ are different undetermined parameters. The two corresponding equations for $Q_{2,1}$ and $Q_{2,-2}$ are obtained by varying $P_{2,1}$ and $P_{2,-2}$ independently.

If we put $P_{n,l} = P_{n,-l-1}$, $Q_{n,l} = Q_{n,-l-1}$ as in the non-relativistic treatment, equations (57) and (58) become the same except for the second term and the terms arising from interactions within the (2*p*) group. These terms are outlined in the equations. The relativistic self-consistent field equations without exchange are derived from (57) and (58) by omitting terms in Q's other than $Q_{2,1}$ and $Q_{2,-2}$.

In conclusion, it is a pleasure to record my thanks to Professor D. R. Hartree, F.R.S., for suggesting the possibility of this extension of the self-consistent field method, and for his interest and criticism during the progress of the work.

SUMMARY

A relativistic treatment of a many-electron atom on the lines of the self-consistent field method with " exchange," is developed. Magnetic interactions and retardation effects are, for the present, neglected. Tables are given from which the expression for the total energy of an atom containing *s*, *p*, *d* electrons may be constructed. An example is given of the derivation of the relativistic self-consistent field equations, including " exchange " terms, by means of a variation principle.

Phase-Equilibrium of Supraconductors in a Magnetic Field

By H. LONDON, Clarendon Laboratory, Oxford

(*Communicated by F. A. Lindemann, F.R.S.—Received June* 13, 1935)

At present it seems to be doubtful whether in the theory of supraconductivity the conception of two phases, a supraconducting and a normal phase, can describe the transformation between the two states and how far the equations which were first proposed for the pure supraconducting state in small fields* remain true, when both phases are present. Especially it seems to be impossible to describe a phase-transition when an electric field is present, as probably occurs when one cools a wire carrying a constant current. However, where there is only a *magnetic* field the problem can be treated consistently, and the results are in agreement with experiment, and so it may be worth while to try the two-phase hypothesis, excluding equilibrium states in an electric field.

The disturbance of supraconductivity by a magnetic field H_T at a temperature T below the transition point has been treated thermodynamically by Gorter† as a phase transition, the supraconducting phase being characterized by a zero value of the magnetic induction B.

If $\Delta F = F_n - F_s$ is the difference in the free energy of the two phases which is determined by the difference of the specific heats, then

$$\tfrac{1}{2}H_T^2V = \Delta F \equiv F_n - F_s, \tag{1}$$

where H_T is measured in rational units and F is not the energy-density but the free energy of the volume V itself.

Here we shall abandon the discontinuous transition from $B = H$ in the normal conductor into $B = O$ in the supraconductor‡ and describe

---

* F. and H. London, 'Proc. Roy. Soc.,' A, vol. 149, p. 71 (1935), referred to as A; 'Physica,' vol. 2, p. 341 (1935).

† 'Nature,' vol. 132, p. 931 (1933); Gorter and Casimir, 'Physica,' vol. 1, p. 306 (1934).

‡ The need to correct the thermodynamic considerations for supraconductors of small dimensions has been remarked already by Gorter in a note published since this paper was written, who has used the resulting deviations to explain the anomalous behaviour of alloys (' Physica,' vol. 2, p. 449 (1935)).

In the present paper an exact derivation of the equilibrium condition will be given instead the connexion between magnetic field and current density by the new fundamental equation :*

$$\Lambda c \operatorname{curl} \mathbf{J} = -\mathbf{H} \qquad \mathbf{B} = \mathbf{H} \tag{2A}$$

where

$$\Lambda = m/ne^2$$

is a specific constant of the material which depends on the number n of the electrons which carry the supraconducting current.

If we neglect the displacement current, so that $c \operatorname{curl} \mathbf{H} = \mathbf{J}$, it follows from these equations that H and J decrease exponentially as the distance from the surface increases, so that at a depth

$$\mathrm{D} = \sqrt{\Lambda c^2}$$

they have the *e*th part of their surface values. Assuming for n the number of valency electrons per cubic centimetre we find for D the order of magnitude 10^{-6} cm; assuming only the number of those electrons which are not yet in their lowest state at the transition temperature we get for D about 10^{-4} cm.

As to the behaviour in an electric field several possibilities have been discussed.† Here we shall assume the simplest one,

$$\Lambda \dot{\mathbf{J}} = \mathbf{E}, \tag{2B}$$

into which the other formulations degenerate when we have no external electric fields and consider only slow variations, as we shall do in the following.

For this system of equations the law of conservation of energy can be derived in a similar way, as in the previous paper,‡ and one gets

$$\frac{\partial \mathrm{W}}{\partial t} + \operatorname{div} \mathbf{S} = 0,$$

where the energy density

$$\mathrm{W} = \tfrac{1}{2}(\mathbf{E}^2 + \mathbf{H}^2 + \Lambda \mathbf{J}^2), \tag{3}$$

and the energy flow

$$\mathbf{S} = c\,[\mathbf{EH}].$$

---

which shows that the kinetic energy of the current plays an essential role and leads to some new conclusions with regard to pure supraconducting metals. I am indebted to Dr Gorter for giving me his manuscript before publication.

* A, p. 73, equation (6).

† 'Physica,' vol. 2, p. 351 *ff.* (1935); v. Laue, F. and H. London, 'Z. Physik,' vol. 96, p. 363 (1935).

‡ A, p. 78.

The stress tensor has the form

$$T_{ik} = E_iE_k - \tfrac{1}{2}\delta_{ik}E^2 + H_iH_k - \tfrac{1}{2}\delta_{ik}H^2 - \Lambda(J_iJ_k - \tfrac{1}{2}\delta_{ik}J^2), \qquad (4)$$

with

$$\delta_{ik} = \begin{cases} 0 \text{ for } i \neq k \\ 1 \text{ for } i = k, \end{cases}$$

i, k each of them denoting the co-ordinates x, y, z, and for stationary conditions

$$\text{Div}\, T = 0. \qquad (4\text{A})$$

The tangential components of the field strengths E and H are continuous at the surface, but the current density J is discontinuous. Outside J is zero in the absence of electric fields, whereas in the supraconductor very high values are possible as the current is confined to the thin surface layer. For supraconductors whose dimensions are large compared with D and whose surface has only a small curvature compared with 1/D,

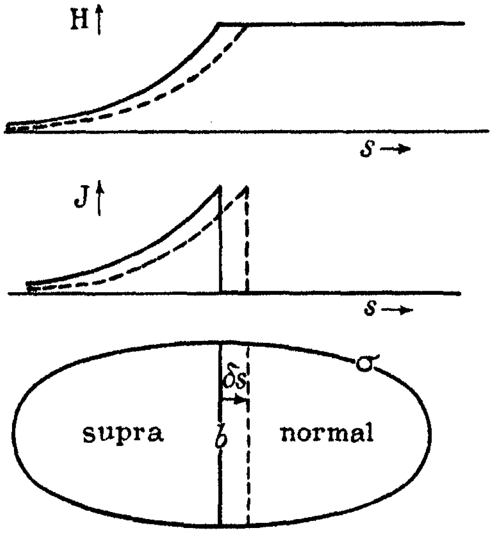


FIG. 1

$$\Lambda J^2 \equiv \left(\frac{D}{c}\right)^2 J^2 = H^2. \qquad (5)$$

1—THE EQUILIBRIUM CONDITIONS

Let us consider a space V partly supraconducting, partly normal, the boundary surface between the two phases having the area b, fig. 1. A magnetic field has to be applied which makes possible the co-existence of the two phases. The diagrams show how this field and the screening currents vary along a line s perpendicular to the boundary surface. The dimensions of the space may be so small that we can regard b as a plane and all fields and the current density as constant within it. Let the boundary surface be subject to a virtual displacement δs in the direction of the normal, a positive δs denoting expansion of the supraconducting phase. If the velocity of the displacement is v the corresponding time is $\delta t = \frac{1}{v}\,\delta s$.

When the boundary surface is moving, rotational electric fields are induced. In contrast to the excluded stationary electric fields they can be described by the equations (2), which include Maxwell's induction law,

$$c \operatorname{curl} \mathbf{E} = -\dot{\mathbf{B}}. \tag{6}$$

The two-phase hypothesis implies the fact that in the interior of either phase only electrical processes governed by (2) or Ohm's law occur, even for a moving boundary surface.

Therefore we can employ the law of conservation of energy for any volume-element which lies entirely in one phase

$$\frac{\partial W}{\partial t} + \operatorname{div} \mathbf{S} + Q = 0,$$

where W is the energy density, $\mathbf{S}$ the energy flow, and Q the Joule-heat.

If we make v small enough, we can neglect the Joule-heat and all other terms where E has a power higher than one, since E is proportional to v (according to equation (6)).

Then we have

in the supraconductor (equation (3))	in the normal conductor
$W = \frac{1}{2}(\mathbf{H}^2 + \Lambda \mathbf{J}^2)$	$W = \frac{1}{2}\mathbf{H}^2$
$S = c\,[\mathbf{EH}]$	$S = c\,[\mathbf{EH}]$
$Q = 0$	$Q = 0$

As we can divide our space at any given time into volume elements which lie only in one phase, we can integrate over the whole volume V and transform the integral of the divergence into an integral over the surface σ surrounding the whole volume,

$$\int_V \frac{\partial W}{\partial t}\, dV = -\int_\sigma S_n\, d\sigma. \tag{7}$$

As we wish to write the differentiation-sign before the integral we have to consider the discontinuity of W at the boundary surface of the two phases, since it is moving with the velocity v. The discontinuity amounts to $\frac{1}{2}\Lambda J_b^2$, J_b denoting the current density at the boundary surface (on its supraconducting side). The supraconducting part of the integration volume V_s increases according to $dV_s/dt = vb$, therefore we have to write

$$\frac{\partial}{\partial t}\int W dV = \int \frac{\partial W}{\partial t} dV + vb\tfrac{1}{2}\Lambda J_b^2.$$

Applying this together with (7) to our displacement we obtain

$$\delta \int W dV = -\delta t \int S_n \, d\sigma + \delta s b \tfrac{1}{2}\Lambda J_b^2.$$

This equation tells us that when the supraconducting region is expanding, the increase of the electric energy is higher by the amount $\delta s b \frac{1}{2}\Lambda J_b^2$ than the energy flowing in. The generation of this additional energy is localized just at the boundary surface of the two phases.

At this surface an energy has to disappear which we have not yet introduced in the energy equation. It is due to the phase transformation which occurs here. For an isothermal process we have to use the free energy, which decreases on account of the phase-transformation by the amount

$$\delta \Delta F = \frac{\partial \Delta F}{b \partial s} b \delta s.$$

The change of the *total free energy* in the volume V is represented by this quantity taken with a minus sign, plus the change of the electric energy:

$$-\delta \Delta F + \delta \int W dV,$$

while the *work* introduced from outside is given by:

$$\delta A = -\delta t \int S_n \, d\sigma.$$

A reversible phase-transformation can only take place if

$$-\delta \Delta F + \delta \int W dV - \delta A = 0.$$

Thus, substituting from the preceding formulae, we obtain the following equilibrium condition:

$$\frac{\partial \Delta F}{b \partial s} = \tfrac{1}{2}\Lambda J_T^2 \equiv \tfrac{1}{2}\left(\frac{D}{c}\right)^2 J_T^2. \tag{8}$$

Instead of a magnetic threshold value we have now a threshold value J_T for the *current density* J_b at the surface.

For supraconductors of large dimensions and small curvature, magnetic fields and current densities are in fixed relation according to equation (5) and the free energy is certainly proportional to the volume, so that $\frac{\partial \Delta F}{b \partial s} = \frac{\Delta F}{V}$. In this case from (6) we again obtain Gorter's relation

$$\tag{1} \tfrac{1}{2} H_T^2 V = \Delta F,$$

which is proved by experiments with pure metals.*

For supraconductors or supraconducting phases of small dimensions, however, we must expect deviations from this law; this we shall investigate in § 2.

The result expressed by equation (6) can easily be understood intuitively. From (2) it follows that a volume element which becomes supraconducting in a magnetic field is provided with a certain current in the very moment of its transformation, and, from the reversibility of this process, we can conclude that in the converse case also the current-carrying supraconductor changes directly into the normal state with no current, and the current does not die out afterwards producing Joule heat.† The rise in the energy of a supraconductor due to the current amounts to $\tfrac{1}{2} \Lambda J_b^2$. The additional magnetic energy does not change when a small volume element is transformed. If $J_b = J_T$, the total free energy of the supraconductor just reaches the level of the normal phase; when this level is exceeded the transformation into the normal state takes place. The generation of current in the part which becomes supraconductive takes place without any assistance of an electric field and is only due to forces which come from the decrease of the free energy caused by the phase-transformation. All other effects necessarily connected with the movement of the boundary surface, namely, the decrease of H and J behind the surface, occur automatically for electrodynamic reasons, just in the same way as if the current in the part which actually becomes supraconductive were produced by an external circuit.

---

* Keesom and Kok, 'Physica,' vol. 1, pp. 503, 595 (1935).

† In this respect equations (2) differ in principle from the formulation proposed by Becker, Heller, and Sauter ('Z. Physik,' vol. 85, p. 772 (1933)) before the reversibility of the phase-transformation was known. According to their formulation at the transition into the supraconducting state the magnetic field would remain unchanged and no external work would be done, while at the disturbance of the supraconductivity the energy of any induced current would be transformed into heat. Therefore no change in the free energy would be possible.

Another derivation of (8) can be obtained by means of the stress tensor, so, as in fig. 1, with $B = 0$ the work done by the supraconductor at a phase-transformation can be considered to be done against the Maxwell stresses $\frac{1}{2}H^2$.* In our case the stress tensor (4) of the supraconducting phase has terms containing the current density. As these do not exist in the tensor of the normal phase the stress has a discontinuity at the boundary; in our case, where we have only the tangential component $J_t = J_b$, the discontinuity amounts to $-\frac{1}{2}\Lambda J_b^2$ and gives rise to a surface force of this magnitude in the direction of the normal. This is the only force produced by the stresses, as H has not such a discontinuity; inside the supraconductor $\operatorname{Div} T = 0$ according to (4A) and therefore no volume force occurs. Thus the work to be done is again $b\delta s\frac{1}{2}\Lambda J_b^2$. Here also we see that the work is localized at the boundary surface, though the decrease of the fields takes place in a layer of finite thickness.

2—SUPRACONDUCTORS OF SMALL DIMENSIONS

Consider a plane supraconducting lamina of the constant thickness $2d$ comparable with the depth of penetration D. Let the y- and z-axis of a rectangular co-ordinate system lie in the plane of the lamina and the origin of x in the middle, so that the surface obeys the equation $x = \pm d$. Let all the magnetic fields point in the z-direction (longitudinal case) and the currents in the y-direction.

Then we have

$$\operatorname{curl} \mathbf{H} = \operatorname{curl}_y \mathbf{H} = -\,\partial H_z/\partial x;\ \operatorname{curl} \mathbf{J} = \operatorname{curl}_z \mathbf{J} = +\,\partial J_y/\partial x,$$

and equation (2) gives for stationary conditions (where $c \operatorname{curl} \mathbf{H} = J$):

$$H = \bar{H}\,\frac{\cosh x/D}{\cosh d/D} + \frac{1}{2}\,\frac{I}{c}\,\frac{\sinh x/D}{\sinh d/D}$$

$$J = \frac{c}{D}\left(\bar{H}\,\frac{\sinh x/D}{\cosh d/D} + \frac{1}{2}\,\frac{I}{c}\,\frac{\cosh x/D}{\sinh d/D}\right).$$

Here the integration constants $\bar{H} = \frac{1}{2}\{H(d) + H(-d)\}$ and $I = c\{H(d) - H(-d)\}$ are determined in such a way that they denote the arithmetical means of the field-strengths at the two surfaces of the lamina and the total current per centimetre flowing through it. For the densities J_+ and J_- of the current flowing at the two surfaces $x = \pm d$ we have

$$\frac{D}{c}\,J_\pm = \pm\,\bar{H}\,\operatorname{tangh}\frac{d}{D} + \frac{1}{2}\,\frac{I}{c}\,\operatorname{ctgh}\frac{d}{D}\,. \tag{9}$$

---

* *See* Gorter, 'Physica,' vol. 2, p. 450, footnote 2, referring to a remark of R. Becker.

The current density is always a maximum at one of the surfaces (or at both), so we have as the condition that the lamina is supraconducting

$$\left.\begin{matrix}|J_+| \\ |J_-|\end{matrix}\right\} \leqq J_T. \tag{10}$$

In order to demonstrate clearly the deviation which is caused by the equilibrium condition (8) in contrast to that resulting from the description $B = 0$, we shall introduce the magnetic field

$$H_T = \frac{D}{c} J_T = \sqrt{\frac{2\partial\Delta F}{b\partial s}}. \tag{11}$$

It should be emphasized that H_T is only an abbreviation for the thermodynamical quantity at the right-hand side of (11) and is *not* to be interpreted as a threshold value.

Then it follows from (9) and (10)

$$\left| \pm \bar{H} \operatorname{tangh} \frac{d}{D} + \tfrac{1}{2} \frac{I}{c} \operatorname{ctgh} \frac{d}{D} \right| \leqq H_T$$

or

$$|\bar{H}| \operatorname{tangh} \frac{d}{D} + \tfrac{1}{2} \left|\frac{I}{c}\right| \operatorname{ctgh} \frac{d}{D} \leqq H_T. \tag{12}$$

We consider the following special cases:

1—The lamina may be placed in a homogeneous external field H_1 with no current led into it from outside. Then we have $\bar{H} = H_1$ and we get from (12) for H_1 the threshold value

$$H_{1T} = H_T \operatorname{ctgh} \frac{d}{D}, \tag{13}$$

which is clearly larger than H_T.

2—There may be no external field but a current I led in from outside which produces at the surface the magnetic field $H_2 = \pm \tfrac{1}{2} I/c$.

This field must be smaller than

$$H_{2T} = H_T \operatorname{tangh} \frac{d}{D}. \tag{14}$$

The maximum currents are therefore smaller than those following from $B = 0$, and instead of Silsbee's relation $H_{1T} = H_{2T}$ we obtain*

$$H_{1T} = H_{2T} \operatorname{tangh}^2 \frac{d}{D}. \tag{15}$$

---

* By equations (14) and (15) only the lowest point of the transition curve (giving the resistance as a function of temperature or of current) is defined, where no electric field is present.

Generally the inequality (12) defines in the ($\bar{H}$, I)-plane a rhombus, the interior of which contains all combinations of $\bar{H}$ and I for which supraconductivity can persist, fig. 2. The points 1 and 2 correspond to the two special cases given.

Similar equations hold for thin supraconducting wires of the radius R:

1—Wire in a longitudinal field without current

$$H_{1T} = H_T \frac{iJ_0(iR/D)}{J_1(iR/D)}. \tag{13'}$$

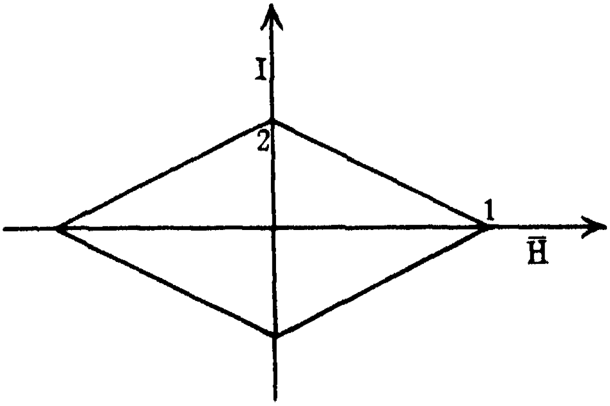


FIG. 2—Relation between threshold current and threshold field.

2—Wire with current I producing the field $H_2 = \frac{I}{2\pi c R}$

$$\frac{I_T}{2\pi c R} = H_{2T} = H_T \frac{J_1(iR/D)}{iJ_0(iR/D)}, \tag{14'}$$

where J_0 and J_1 are Bessel functions.

The combinations of H_1 and I for which supraconductivity can persist are in this case confined to an ellipse with the axes H_{1T} and I_T.

3—THERMODYNAMIC CONCLUSIONS

From (13) one is tempted to conclude that every supraconductor in a magnetic field above H_T should split up into a great number of thin supraconducting laminae or fibres separated from each other by thin, normal conducting regions, as then supraconductivity could persist at these higher fields. Macroscopically we should observe the disappearance, or at any rate reduction, of the resistance and of the induction B, and a free energy (determined by calorimetric measurements) less than that of the normal conductor. In respect to the resistance, experiment shows that this is not true for undeformed single crystals of pure metals*; for the two other phenomena definite results with single crystals are not yet available with regard to the thermal and magnetic deviations from the normal state found by Keesom and Kok,† and by Keeley, Mendelssohn, and Moore, and Babbitt‡ with polycrystalline material just above the threshold-field. The various experiments§ show, however, that these deviations are confined to a small region, so that for the time being it seems to be reasonable to assume that the ideal behaviour is the discontinuous disappearance of all supraconduction phenomena above the threshold-field. This means

$$H_{1T} \leqq H_{T\infty} = \sqrt{2\left(\frac{\Delta F}{V}\right)_{\propto}}, \tag{16}$$

where $H_{T\infty}$ is the magnetic threshold value of large supraconductors, for which equation (1) holds.

We are therefore forced to conclude that the thermodynamic quantity $H_T = \sqrt{2\frac{\partial \Delta F}{b\,\partial s}}$ in (13) has not the constant value $H_{T\infty}$, but must decrease with the dimensions. In other words, the splitting-off must be connected with an increase of free energy $F_s = F_n - \Delta F$.|| Using (13) and (11) we get from (16) the conditions:

for a lamina,

$$\frac{\partial \Delta F(d)}{b\,\partial d} \leqq \left(\frac{\Delta F}{V}\right)_{\infty} \operatorname{tangh}^2 \frac{d}{D}, \tag{17}$$

for a wire,

$$\frac{\partial \Delta F(R)}{2\pi R l\,\partial R} \leqq \left(\frac{\Delta F}{V}\right)_{\infty} \left(\frac{J_1(iR/D)}{iJ_0(iR/D)}\right)^2, \tag{17'}$$

where we have substituted in (17′) $b = 2\pi R l$, l denoting the length of the wire.

---

* 'Comm. Phys. Lab., Leiden,' No. 214c.

† Keesom and Kok, 'Physica,' vol. 1, p. 503 (1934).

‡ Keeley, Mendelssohn, and Moore, 'Nature,' vol. 134, p. 773 (1934); Mendelssohn and Babbitt, 'Proc. Roy. Soc.,' A, vol. 151, p. 332 (1935).

§ Meissner and Ochsenfeld, 'Naturwiss.,' vol. 21, p. 787 (1933); 'Phys. Z.,' vol. 35, p. 954 (1934); de Haas and Casimir, 'Physica,' vol. 1, p 291 (1934); Mendelssohn and Babbitt, 'Nature,' vol. 133, p. 413 (1934); *see also* Rjalinin and Shubnikow, 'Phys. Z. Sowjet,' vol. 5, p. 641 (1934); 'Nature,' vol. 134, p. 286 (1934); 'Phys. Z. Sowjet,' vol. 6, p. 557 (1934); 'Nature,' vol. 135, p. 109 (1935); Tarr and Wilhelm, 'Canad. J. Res.,' vol. 12, p. 265 (1935).

|| In his new paper ('Physica,' vol. 2, p. 449 (1935)) Gorter comes to the more special assumption of a lower limit for the size of a supraconducting phase.

Integrating we obtain

$$\Delta F(d) \leqq \left(\frac{\Delta F}{V}\right)_\infty bd\left(1 - \frac{D}{d}\tanh\frac{d}{D}\right) \tag{18}$$

$$\Delta F(R) \leqq \left(\frac{\Delta F}{V}\right)_\infty \pi R^2 l\left(1 - \frac{D}{R}\frac{2J_1(iR/D)}{iJ_0(iR/D)}\right). \tag{18'}$$

Here we have determined the integration constant in such a way that $\Delta F \leqq 0$ for $d = 0$ or $R = 0$.*

If d, $R \gg D$ we can write

$$F_n - F_s \equiv \Delta F(d) \leqq \left(\frac{\Delta F}{V}\right)_\infty b(2d - 2D)$$

$$F_n - F_s \equiv \Delta F(R) \leqq \left(\frac{\Delta F}{V}\right)_\infty l(\pi R^2 - 2\pi RD).$$

Here the free energy difference becomes proportional to a volume which can be obtained from the true volume by subtracting a layer of the thickness D all over the surface. We can therefore speak of a surface energy or attribute to the supraconducting phase with respect to the normal phase a surface tension

$$\gamma \geqq D\left(\frac{\Delta F}{V}\right)_\infty . \tag{19}$$

This surface tension will, however, produce an observable effect only if the larger-sign holds in (19). If the equality-sign is true the pressure produced by the surface tension on a curved surface would just be compensated by a decrease of the electromagnetic stress; for the physical meaning of the equality-sign is that at a given temperature supraconductors of any size and curvature are in equilibrium with the normal phase in the same magnetic field.

Equations (17) and (18) are not meant also to apply to thin films or wires. Here we cannot predict anything about H_T. But from the deviations from Silsbee's relation the penetration depth D could be calculated and then H_T could be determined by (13) or (14). However, there are very great experimental difficulties in realizing ideal conditions. Polycrystalline structure and internal stresses (caused, for example, by the different thermal expansion of the supraconductor and the supporting material) render the behaviour of pure metals similar to that of alloys.†

---

* Equation (18) is simply a *deduction* from (16), therefore in contrast to (17) only a necessary and not a sufficient condition for preventing the splitting-off.

† Mendelssohn, "Discussion on Supraconductivity," 'Proc. Roy. Soc.,' A, vol. 152, p. 1 (1935).

Irregularities in thickness would cause a failure of Silsbee's relation independently of equation (15), as has already been pointed out in the first investigation of thin films by Sizoo and Kamerlingh Onnes.[*] Thus it seems to be premature to apply our formulae directly to recent measurements in Toronto,[†] although they show a qualitative agreement assuming a penetration depth of the order of 10^{-4} cm.

By taking the larger-sign in (17) (where $\partial \Delta F\,(d)/b\partial d < (\Delta F/V)_{\infty}$ may still hold) the behaviour of the alloys[‡] can be explained. Here a high magnetic threshold value for electric conductivity[§] and no expulsion of magnetic flux[||] has been observed, the free energy difference being small compared with the threshold value.[¶] The inhomogeneous structure of the alloys may perhaps facilitate the formation of small supraconducting regions. Several explanations can be given for experiments which show that relatively small magnetic fields applied after cooling penetrate the alloy and for the observation[**] that for alloys the current threshold value is much smaller than it should be according to Silsbee's relation. One could, for instance, make suitable assumptions relating *d* or R to the field-strength, so that one always is near to point 1 in fig. 2, or ascribe these effects to a net- or sponge-structure of the supraconducting phase such as has been made probable experimentally by Mendelssohn.[††] Finally, one could suppose that the normal conducting separating surfaces may lie perpendicular to the direction of the current (in the case of fibres, indeed, they must do so). There is too vast a field open for arbitrary assumptions for it to be worth while going into details before more experimental data are available.

[*] 'Comm. Phys. Lab., Leiden,' No. 180a.

[†] Burton, 'Nature,' vol. 133, p. 459 (1934); Burton, Wilhelm, and Misener, 'Trans. Roy. Soc. Canad.,' (3), vol. 28, p. 65 (1934); Misener and Wilhelm, 'Comm. McLennan Lab. Toronto,' No. 72; Misener, Grayson Smith, and Wilhelm, *ibid.*, No. 78 (1935).

[‡] This is analogous to the assumption, made in the new paper of Gorter ('Physica,' vol. 2, p. 449 (1935)), of a small lower limit for supraconducting regions in alloys.

[§] de Haas and Voogd, 'Comm. Phys. Lab., Leiden,' No. 208b.

[||] Keeley, Mendelssohn, and Moore, 'Nature,' vol. 134, p. 773 (1934); Tarr and Wilhelm, 'Canad. J. Res.,' vol. 12, p. 265 (1935); Rjalinin and Shubnikow, 'Nature,' vol. 135, p. 582 (1935); 'Phys. Z. Sowjet,' vol. 7, p. 124 (1935); Mendelssohn and Moore, 'Nature,' vol. 135, p. 826 (1935).

[¶] Mendelssohn and Moore, 'Nature,' vol. 134, p. 773 (1934); 'Proc. Roy. Soc.,' A, vol. 151, p. 334 (1935); Shubnikow and Chotkewitch, 'Phys. Z. Sowjet,' vol. 6, p. 605 (1934).

[**] De Haas and Casimir, 'Nature,' vol. 135, p. 30 (1935); Keesom, 'Physica,' vol. 2, p. 35 (1935); Rjalinin and Shubnikow, Mendelssohn and Moore. (*See* footnote ||.)

[††] *See* footnote †, p. 660.

A smaller-sign in (17) could give an explanation for the hysteresis, for then the formation of the first supraconducting nucleus in a magnetic field would require a higher amount of energy than is necessary for further expansion. In this case the hysteresis would be analogous to the super-cooling of a liquid or the super-saturation of a vapour. The hysteresis phenomena we are considering here are those that have been observed in resistance measurements in a magnetic field, the measuring current being commutated after each measurement.* Here the occurrence of the supraconductive state is delayed in an unreproducible manner, *i.e.*, the resistance of the wire remains at the normal value after the magnetic field is decreased below the threshold while the disturbance of supraconductivity takes place always at the same field given by (1) which determines the field for a reversible transition. Therefore it is the *normal* state which represents the unstable part of this hysteresis phenomenon.

There is still another sort of hysteresis. It is due to the fact that a supraconducting *ring* prevents the magnetic lines of force from crossing it. Therefore if the supraconducting phase forms a ring, the system can come into a metastable state from which it can go irreversibly into a more stable one if the magnetic field rises above the threshold value. Such effects have often been observed in the course of the last year during experiments on the expulsion of magnetic field,† definite hysteresis loops being obtained if the magnetic flux through the specimen was plotted against the external magnetic field. This phenomenon, which might be called " hysteresis of the flux," can only occur when a *supraconducting* phase is present; no more than one *point* of the loop can belong to a state in which the *whole* specimen is normal conducting. By this it can be distinguished from the " hysteresis of phase-transition " considered above, which is characterized by *normal conductivity, i.e.*, normal resistance and normal magnetic flux in a region below the thermodynamic threshold value.

In conclusion, I should like to express my thanks to Professor F. A. Lindemann, F.R.S., for his kind hospitality at the Clarendon Laboratory and his interest in my work.

I am indebted to Dr. K. Mendelssohn for the communication of several

---

* 'Comm. Phys. Lab., Leiden,' No. 212c.

† Mendelssohn and Babbitt, 'Nature,' vol. 133, p. 413 (1934); 'Proc. Roy. Soc.,' A, vol. 151, p. 332 (1935); *see also* Rjalinin and Shubnikow, 'Phys. Z. Sowjet,' vol. 5, p. 641 (1934); 'Nature,' vol. 134, p. 286 (1934); 'Phys. Z. Sowjet,' vol. 6, p. 557 (1934); 'Nature,' vol. 135, p. 109 (1935); Tarr and Wilhelm, 'Canad. J. Res.,' vol. 12, p. 265 (1935).

unpublished details about his researches in this laboratory, and to Mr. B. V. Rollin for his kind help in the translation of this paper.

SUMMARY

The disturbance of supraconductivity by a magnetic field is treated thermodynamically from the electrodynamic equations of the supraconductor given recently by F. and H. London. It follows that a threshold value of the current density and not of the magnetic field is the decisive quantity.

Application to supraconducting regions of small dimensions yields a deviation from Silsbee's relation between magnetic and current-threshold value and leads to the conclusion that the free energy difference between the normal and the supraconducting phase must depend on the dimensions of the supraconductor. This may be interpreted as a surface tension of the supraconducting phase and suggests an explanation of the hysteresis.

The Nuclear Spin of Iodine

II—Fine Structure in the Arc Spectrum and a Fine Structure Perturbation Effect

By S. TOLANSKY, Ph.D., Manchester University

(*Communicated by W. L. Bragg, F.R.S.—Received June* 11, 1935)

INTRODUCTION

The arc spectrum of iodine was first examined for fine structure by Wood and Kimura,* who failed to detect structure in any of the lines. The failure was probably due to the use of an unsatisfactory source and to the smallness of the resolving powers employed. The author then examined the iodine arc spectrum that was emitted by a high frequency electrodeless discharge in pure iodine vapour,† using a variable gap Fabry-Perot interferometer as the resolving instrument. Structure was observed in eleven lines and an attempt at analysis was made. The nuclear spin of iodine was then unknown and, as only partial resolution

* 'Astrophys. J.,' vol. 48, p. 181 (1917).

† 'Proc. Roy. Soc.,' A, vol. 136, p. 585 (1932).

was achieved, interpretation proved to be very difficult and ambiguous. Furthermore, at that time the existence of fine structure perturbations was not yet even suspected and, as will be shown later, one of the important *s* terms is perturbed.

The nuclear spin of iodine was determined* from extensive measurements of the fine structures in the lines of the first spark spectrum. The lines produced in a water-cooled hollow cathode were investigated and, owing both to the sharpness of the individual components and to the relatively wider structures associated with the ionized atom, an unambiguous value for the spin was found (5/2). Knowing the spin with certainty, this can now be applied to the interpretation of the structures in the arc lines. However, realizing the superiority of the hollow cathode as a source, a complete re-examination of the spectrum was made. This revealed the fact that the previous measurements were all approximately correct, the differences being accounted for by the higher resolution now attained.

The gross structure multiplet analysis of the arc spectrum is very incomplete. A number of attempts has been made at analysis, the most complete appearing to be that of Evans.† About 150 lines were measured, but only seven terms were completely identified, although values were found for ten others and the numerical positions of another twenty unknown terms ascertained. In the main the fine structure analysis supports the gross structure term allocations made by Evans, which are therefore used here. The line transitions examined for structure are shown in fig. 1. All the lines involve either $6s\ {}^4P_{5/2}$ or $6s\ {}^4P_{3/2}$ as lower terms, and as in addition four of the upper *p* electron terms are each involved twice, opportunity is afforded for a check on the analysis made. On the basis of White's vector coupling‡ scheme, it is to be expected that the fine structure interval factor for $6s\ {}^4P_{5/2}$ should be greater than that of $6s\ {}^4P_{3/2}$, whilst by analogy with the author's observation upon the Br I spectrum,§ some of the upper *p* electron terms may be expected to show moderately large interval factors.

EXPERIMENTAL

The apparatus used was similar to that described in Part I. The arc spectrum was excited in a water-cooled hollow cathode containing iodine and only a trace of helium. The spectrum could also be excited if hydrogen or nitrogen were used as carriers instead of helium, which has a strong tendency to bring out the spark spectrum only. The structures were examined with a large aperture Fabry-Perot interferometer crossed with a two-prism glass spectrograph with camera focal length equal to 128 cm. Plate separations between 5–25 mm were used. When employing Ilford hypersensitive plates exposure times varied from a few minutes to one hour.

---

* Part I, 'Proc. Roy. Soc.,' A, vol. 149, p. 269 (1935).

† 'Proc. Roy. Soc.,' A, vol. 133, p. 417 (1931).

‡ 'Phys. Rev.,' vol. 35, p. 1146 (1930).

§ 'Proc. Roy. Soc.,' A, vol. 136, p. 585 (1932).

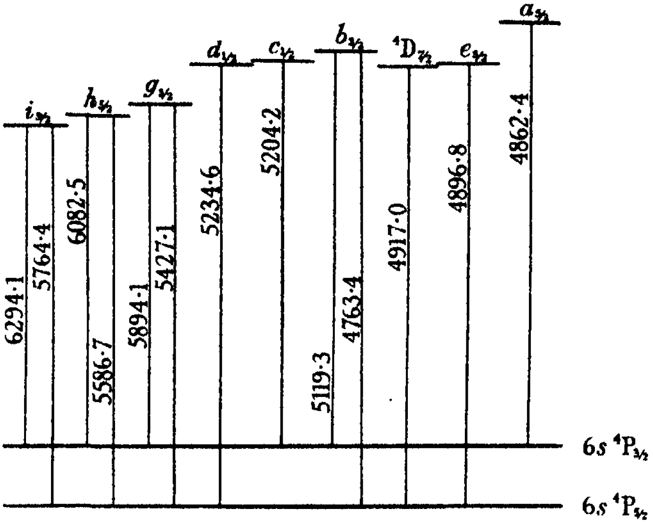


FIG. 1—Line transitions examined for fine structure.

Table I gives a list of the fine structures observed, in units of $\text{cm}^{-1} \times 10^{-3}$, and below each component the visually-estimated intensity is given in a bracket. The line allocations are those given by Evans. The degree of sharpness of the components varies considerably, *i.e.*, some are complex, and this fact was taken into consideration during the analysis.

ANALYSIS

λ 5427·1—From the appearance of the structures it was apparent that in practically every case the upper term has an appreciably large interval factor, the result of which causes excessive overlapping of the fine structure components. The nuclear spin and the J values being known, it is possible to construct Fisher-Goudsmit diagrams* which can be relied upon. These diagrams were used for the analysis of all the lines.

---

* 'Phys. Rev.,' vol. 37, p. 1059 (1931).

The first line selected for analysis was λ 5427·1 (6*s* $^4P_{5/2}$—7*p* $g_{3/2}$). This was chosen because the components were sharp, the resolution was good; the lower term should be wide, and the upper J value being the smallest possible combining with 6*s* $^4P_{5/2}$, the complexity of the pattern is thereby reduced. However, it was found to be totally impossible to fit the observed pattern into a Fisher-Goudsmit diagram, and, in fact, this was true for all the lines involving 6*s* $^4P_{5/2}$. After examining various possibilities, it was concluded that this term must be perturbed by a term with $J = \frac{3}{2}$, the perturbing term lying deeper than 6*s* $^4P_{5/2}$. The perturbation is probably of a similar nature to that found by Schüler and Jones in the 6 $^1D_2$ term of the mercury arc spectrum.* These authors found that two terms perturbed each other when falling close together, the displacements

TABLE I—FINE STRUCTURES OF THE IODINE ARC LINES

Wave-length Angstroms	Classification	Structure, cm^{-1} × 10^{-3}						
			Red				Violet	
6294·1	6*s* $^4P_{3/2}$—7*p* $i_{3/2}$	0	61	124	167			
		(9)	(7)	(5)	(3)			
6082·5	6*s* $^4P_{3/2}$—7*p* $h_{3/2}$	0	86	172	242			
		(4)	(3)	(2)	(1)			
6024·1	— — — —	0	47	131	213	280	374	464
		(2)	(4)	(4)	(2)	(1)	(8)	(2)
5894·1	6*s* $^4P_{3/2}$—7*p* $g_{3/2}$	0	55	110	145			
		(6)	(5)	(3)	(2)			
5586·7	6*s* $^4P_{5/2}$—7*p* $h_{3/2}$	0	238	448	618	739	804	
		(6)	(5)	(4)	(3)	(2)	(1)	
5427·1	6*s* $^4P_{5/2}$—7*p* $g_{3/2}$	0	213	400	537	—	646	
		(5)	(4)	(3)	(2)		(1)	
5234·6	6*s* $^4P_{3/2}$—7*p* $d_{1/2}$	0	47	140				
		(4)	(2)	(3)				
5204·2	6*s* $^4P_{3/2}$—7*p* $c_{1/2}$	0	63	146	189	249		
		(4)	(5)	(7)	(5)	(1½)		
5119·3	6*s* $^4P_{3/2}$—7*p* $b_{3/2}$	0	80	120	174			
		(2)	(3)	(9)	(1)			
4917·0	6*s* $^4P_{5/2}$—7*p* $^4D_{7/2}$	0	100	207	303	388	452	
		(1)	(12)	(10)	(8)	(4)	(5)	
4896·8	6*s* $^4P_{5/2}$—7*p* $e_{3/2}$	0	70	139	191	284	384	
		(10)	(8)	(6)	(4)	(10)	(3)	
4862·4	6*s* $^4P_{3/2}$—7*p* $a_{3/2}$	0	85	203	301	381		
		(1)	(10)	(8)	(6)	(4)		
4763·4	6*s* $^4P_{5/2}$—7*p* $b_{3/2}$	0	100	213	311	380	429	
		(10)	(8)	(6)	(4)	(3)	(3)	

---

* 'Z. Physik,' vol. 77, p. 801 (1932).

in the fine structures being such that there was an apparent repulsion between levels with the same F value. The displacement of each level was proportional to F so far as could be measured. Thus, if the two perturbing terms have identical J values, then the F values are repeated in both, and as a result the interval rule holds both in the perturbed and in the unperturbed structure. If, however, the J values differ, then complementary values cannot be found for each F, so that the perturbation occasions a breakdown in the interval rule.

In $6s\,^4P_{5/2}$ there are six F levels—0, 1, 2, 3, 4, 5 (for I = $\frac{5}{2}$), and as the perturbation is produced by a deeper-lying term with J = $\frac{3}{2}$ (F = 1, 2, 3, 4), the fine structure levels F = 0 and 5 are unaffected whilst the others are displaced. This breaks down the interval rule and explains the failure of a normal Fisher-Goudsmit diagram. The interval rule holds for the F terms 1, 2, 3, 4 and, as it happens, the distortion produced by the non-repulsion of the F value 0 is very small, so that in effect the term contains a regular quintet structure (F = 0, 1, 2, 3, 4), together with a distorted sixth level (F = 5). Thus a modified diagram can be employed and the analysis carried through. The result is shown in fig. 2, from which it is seen that the agreement between the calculated and observed patterns is very good. The dotted level in the lower term illustrates the position F = 5 should occupy for the validity of the interval rule to be retained.

It has been assumed in the analysis of the above line that the interval rule is obeyed in the upper term and, although this is apparently correct, it is true that small deviations could not have been detected because of the complexity of the pattern. Even if they exist, the analysis will not be invalidated. This is true for most of the lines examined. Defining the interval factor as A where the separation between two levels F + 1 and F equals A (F + 1), then the value of A for $6s\,^4P_{5/2}$ is 53·5 and for $7p\,g_{5/2}$ = 10·7. The value 53·5 will be discussed later in relation to the perturbation, and it will be shown that 52 is the unperturbed value.

λ 5894·1—Knowing the value of A for $7p\,g_{5/2}$ the analysis of λ 5894·1 ($6s\,^4P_{3/2}$—$7p\,g_{5/2}$) gives the interval factor for $6s\,^4P_{3/2}$. The structure in this line is very small and the resolution very incomplete. In spite of this, an approximately good agreement was obtained when the interval factor for the lower term was taken to be 24.

λ 5234·6—This line ($6s\,^4P_{3/2}$—$7p\,d_{1/2}$) also exhibits a closely packed pattern, but as the upper J value is $\frac{1}{2}$ the number of components is small, so that a fit is reasonably possible. Taking 24 for the lower term, the value found for $7p\,d_{1/2}$ was 16.

λ 5119·3—$6s\,^4P_{3/2}$—$7p\,b_{3/2}$. This pattern is very narrow and complex because the upper and lower terms have the same J value and interval factors of about the same order, the lower being 24 and the upper 29 approximately.

λ 4763·4—Both this line ($6s\ {}^4P_{5/2}$—$7p\ b_{5/2}$) and the next two afford important checks upon the lower perturbed level. The analysis is shown in fig. 3. Since the line lies right within the strong continuum that is

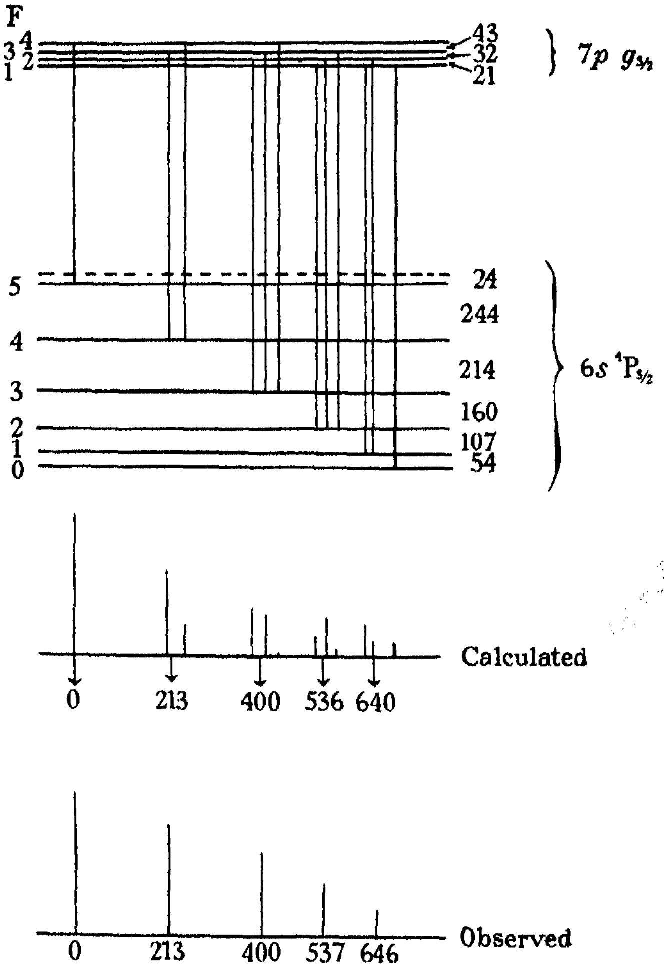


FIG. 2—λ 5427·1 ($6s\ {}^4P_{5/2}$ – $7p\ g_{5/2}$).

associated with the iodine arc spectrum from about λ 4800 to the violet, the weak tail components are extremely difficult to measure, so that, allowing for this, the agreement between the calculated and the observed patterns is very good. The value for A for $7p\ b_{5/2}$ is 34. This is a satisfactory check with the approximate value, 29, given by the previous line, which is very badly resolved.

The above examples illustrate the method of the analysis which was extended to all the lines successfully except λ 4862·4. In particular, λ 4917·0 and λ 4896·8, which both involve $6s\,^4P_{5/2}$, verified the perturbation effect in this term. Table II shows all the observed interval factors, the units being as before cm^{-1} × 10^{-3}.

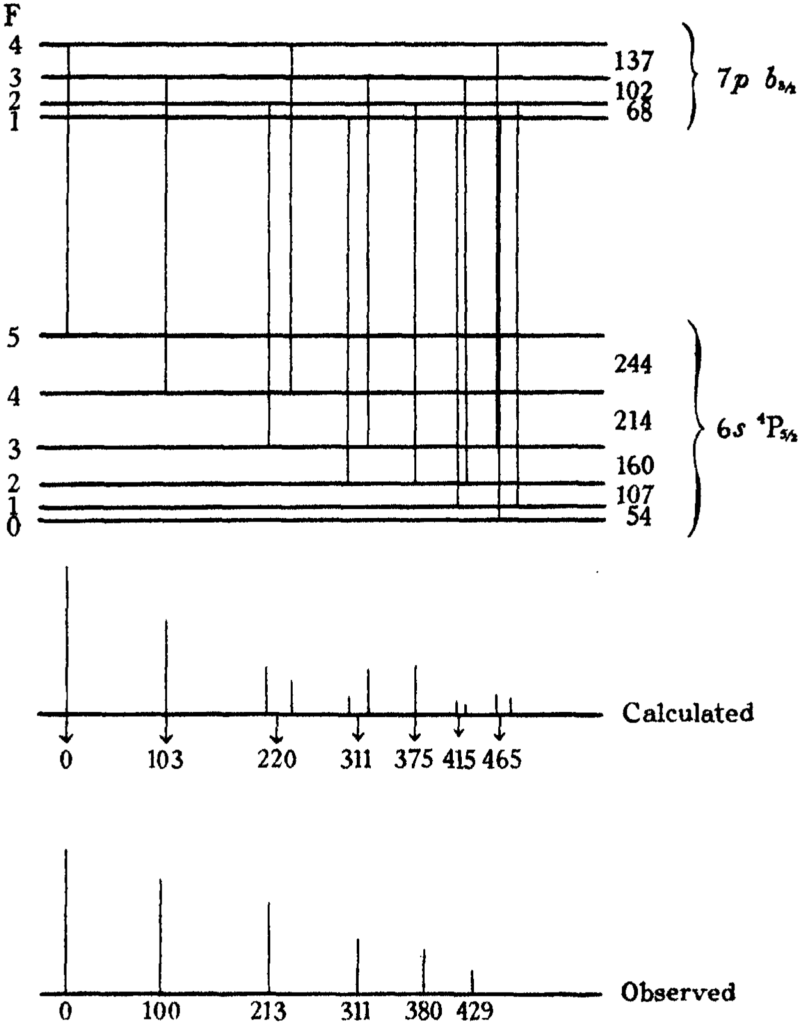


FIG. 3—λ 4763·4 ($6s\,^4P_{5/2}$—$7p\ b_{5/2}$).

It was not found possible to analyse λ 4862·4 ($6s\,^4P_{5/2}$—$7p\ a_{5/2}$) satisfactorily. This may be due to one of two causes: (*a*) the upper term may be strongly perturbed, (*b*) the line may be wrongly classified. Regarding the latter possibility, Iwama,* in his attempt at the gross structure analysis, classifies this line as $6s\ {}^4P_{\frac{3}{2}}$—$7p\ {}^4D_{\frac{7}{2}}$. The question of interpretation remains in doubt.

From figs. 2 and 3 it can be seen that the perturbation effect is determined by the components which are best resolved; hence reliance can be placed upon the perturbation measurement.

TABLE II—FINE STRUCTURE TERM INTERVAL FACTORS IN THE IODINE ARC SPECTRUM

Term	Interval factor
$6s\ {}^4P_{\frac{5}{2}}$	52
$6s\ {}^4P_{\frac{3}{2}}$	24
$7p\ g_{\frac{5}{2}}$	10·7
$7p\ d_{\frac{1}{2}}$	16
$7p\ b_{\frac{3}{2}}$	34
$7p\ h_{\frac{1}{2}}$	–4
$7p\ i_{\frac{5}{2}}$	~7
$7p\ c_{\frac{1}{2}}$	~60
$7p\ {}^4D_{\frac{7}{2}}$	23½
$7p\ e_{\frac{3}{2}}$	48

DISCUSSION

The perturbation of $6s\ {}^4P_{\frac{5}{2}}$ will now be considered. The nature of this is shown in fig. 4. The straight lines on the left show the calculated position of the fine structure levels before perturbation, and the dotted lines to the right the position after perturbation from below by a term with $J = \frac{3}{2}$. Neither $F = 0$ nor $F = 5$ is affected, whilst the other levels are displaced by amounts proportional to F. The hatched line is the position $F = 5$ would occupy if the interval rule held in the perturbed structure. From the displacement of this level the interval factor of the unperturbed structure is found to be 52. The distortion between the levels $F = 0$ and $F = 1$ is so small (about 1½ $cm^{-1} \times 10^{-3}$) that the measurements would fail to show this. As a measure of the degree of perturbation a "perturbation factor" may be conveniently used. This factor will be defined as the displacement per unit F value per unit of interval factor. Thus the $F = 5$ level has to be moved by 24 units for the interval rule to hold. This represents the sum total perturbation displacement for all the levels, *i.e.*, 15 P where P is the perturbation displacement per unit F. If A is the interval factor before perturbation, then the perturbation factor $p = \frac{P}{A}$, *i.e.*, $p = \frac{24}{15 \times 52} = 0 \cdot 03$. This factor is a measure of the degree of perturbation. The same factors, calculated from the data of Schüler and Jones (*loc. cit.*) for mercury, vary between 0·13 and 0·27 for different terms. The degree of perturbation is therefore smaller in the present instance. The perturbing term is also affected, but is unknown. It is probably a term falling accidentally close to $6s\,{}^4P_{5/2}$. It can only be a term with $J = \frac{3}{2}$ for a $J = \frac{1}{2}$ term would only perturb F = 2 and 3, a $J = \frac{5}{2}$ term would produce no distortion and a $J = \frac{7}{2}$ (or more) term would not show a distortion involving the F = 5 level.

---

* 'Sci. Pap. Inst. Phys. Chem. Res. Tokyo,' vol. 15, p. 163 (1931).

In accordance with expectations the interval factor for $6s\,{}^4P_{5/2}$ is greater than that for $6s\,{}^4P_{3/2}$. The values 52 and 24 may be compared with 55 and 42 found for the analogous Br 1 terms. As in the Br 1 spectrum $p^4 . p$, electron terms are capable of producing very wide structures, although not containing a penetrating *s* electron. The interval factors for $7p\ b_{3/2}$, $7p\ e_{5/2}$, and $7p\ c_{1/2}$ are 34, 48, and ∼ 60 respectively. From this it follows that the values of the interval factors of the 6*s* terms are by no means only determined by the penetrating *s* electron, for a 7*p* electron is not likely to contribute to the same extent as a 6*s* electron, and therefore the large values of the interval factors of the *p* terms must be largely due to the $5p^4$ group.

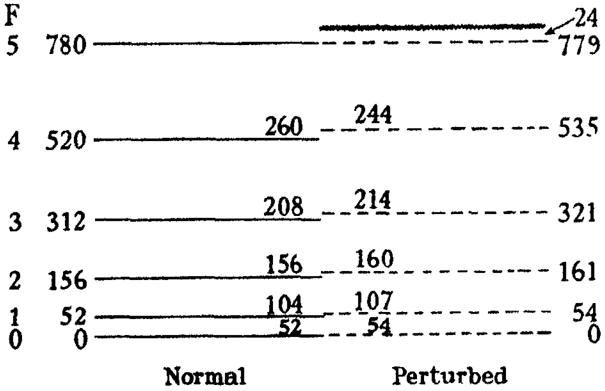


FIG. 4—Perturbation of $6s\,{}^4P_{5/2}$.

In conclusion, it may be noted that the analysis confirms the nuclear spin as $\frac{5}{2}$.

I wish to take this opportunity of expressing my thanks to Professor W. L. Bragg, F.R.S., for his kind interest in this work and for the very excellent experimental facilities he has granted me.

Summary

The fine structures of thirteen of the classified lines of the iodine arc spectrum (region λλ 6300–4760) emitted by a water-cooled hollow cathode discharge have been measured with a silvered Fabry-Perot interferometer and analysed. The nuclear spin value ($\frac{5}{2}$) previously found in the first spark spectrum is confirmed. The 6*s* $^4P_{\frac{3}{2}}$ term is found to be perturbed by a term with a $J = \frac{3}{2}$ causing a breakdown in the fine structure interval rule. Fine structure interval factors are found for ten terms. The interval factor for 6*s* $^4P_{\frac{3}{2}}$ is 2·17 times that of 6*s* $^4P_{\frac{5}{2}}$, but some of the 5*p*^{4} 7*p* terms have fine structure interval factors greater even than that of 6*s* $^4P_{\frac{5}{2}}$. This shows that the *p* electrons make important contributions to the fine structure coupling factors even in the presence of an *s* electron.

The Temperature Dependence of Free Electron Susceptibility

By Edmund C. Stoner, Ph.D. (Cambridge), Reader in Physics at the University of Leeds

(*Communicated by R. Whiddington, F.R.S.—Received June* 15, 1935)

1—Introduction

One of the earliest applications of the Fermi-Dirac statistics was that of Pauli* to the treatment of the paramagnetism, due to the electron spin, of an electron gas. The result he obtained, for low temperatures, may be put in the form

$$M_P = \frac{3}{2} N\mu^2 H/\varepsilon_0, \tag{1.1}$$

where M_P is the total magnetic moment due to the spin effect, N the number of electrons, μ the Bohr magneton, and ε_0 the maximum electron energy in the completely degenerate state. It was later shown by Landau† that electrons, apart from the spin effect, gave a diamagnetic contribution to the susceptibility. The diamagnetic effect (which is zero on a classical basis) arises from the discreteness of the energy states of an electron in a magnetic field. For low temperatures the result obtained is

---

* 'Z. Physik,' vol. 41, p. 81 (1927).

† 'Z. Physik,' vol. 64, p. 629 (1930).

$$M_D = -\tfrac{1}{2} N\mu^2 H/\varepsilon_0, \tag{1.2}$$

where M_D is the diamagnetic contribution to the moment. The spin effect was further considered by Bloch,* who gave, as a higher approximation at low temperatures,

$$M_P = \frac{3}{2}\frac{N\mu^2 H}{\varepsilon_0}\left\{1 - \frac{\pi^2}{12}\left(\frac{kT}{\varepsilon_0}\right)^2\right\}. \tag{1.3}$$

The diamagnetic effect has been further discussed by Teller† and Darwin,‡ mainly with a view to gaining a more detailed physical insight into the origin of the effect; and both the para- and dia-magnetic effects have been treated in some detail in surveys by Van Vleck,§ Pauli,‖ Sommerfeld and Bethe,¶ and Bloch.** At high temperatures the classical formula is obtained for the paramagnetic effect

$$\frac{M_P}{N} = \mu \tanh\left(\frac{\mu H}{kT}\right) = \mu \tanh \beta \doteq \frac{\mu^2 H}{kT}, \tag{1.4}$$

and for the diamagnetic effect

$$\frac{M_D}{N} = -\mu\left(\coth \beta - \frac{1}{\beta}\right) \doteq -\frac{\mu^2 H}{3kT}. \tag{1.5}$$

The character of the variation of the susceptibility with the temperature is to some extent covered by the formulae for the limiting cases given above, but, apart from this, except for Bloch's result (1.3) for the paramagnetism at low temperatures, the variation with temperature has not been explicitly considered in the papers referred to. Recently, a detailed analysis has been made by Niessen,††‡‡ with a view to obtaining the next approximation in powers of T, and also of H, in the low temperature formulae: he considers primarily the joint dia- and para-magnetic effects, and for the dependence on T, obtains

$$\frac{M}{N} = \frac{\mu^2 H}{\varepsilon_0}\left\{1 + \frac{\pi^2}{12}\left(\frac{kT}{\varepsilon_0}\right)^2\right\}. \tag{1.6}$$

---

* 'Z. Physik,' vol. 53, p. 216 (1929).

† 'Z. Physik,' vol. 67, p. 311 (1930).

‡ 'Proc. Camb. Phil. Soc.,' vol. 27, p. 86 (1931).

§ "Theory of Electric and Magnetic Susceptibilities" (Oxford, 1932).

‖ 'Rap. Cons. Phys. Solvay,' vol. 6, p. 175 (1932).

¶ 'Handb. der Physik,' vol. 24 (2), p. 473 (Berlin, 1933).

** 'Handb. der Radiologie,' vol. 6 (2), p. 378 (Leipzig, 1934).

†† *Ibid.*, vol. 1, p. 783 (1934).

‡‡ 'Physica,' vol. 1, p. 979 (1934).

He states further that the bracketed factor is the same for the dia- and para-magnetic effects considered separately. There is thus a discrepancy between the results of Bloch and Niessen for the paramagnetic effect alone, and, further, Niessen's result is very curious, for it would be anticipated that the susceptibility would decrease with increasing temperature, the law of variation passing over continuously to that which holds at high temperatures. Niessen himself refers to the difficulty of giving a physical interpretation of his result.

In view of Niessen's results, it was considered desirable to re-examine the whole question of the temperature dependence of free electron susceptibility, and the primary purpose of this paper is to derive expressions for the temperature variation both at low and at high temperatures. The statistical treatment will be developed by a uniform method for the effects considered separately or together, and in the course of the treatment various formulae, of frequent application in connexion with Fermi-Dirac statistics, are given in what appears to be a more convenient form than those usually employed. Finally, the results will be given in such a form that their numerical significance may be readily appreciated.

Although electrons in metals cannot be considered as free, the results for the susceptibility of free electrons are of considerable importance in connexion with the susceptibility of metals generally. So far as the energy distribution of electrons within a band is of the same type as that for free electrons, the formulae obtained for free electrons will still hold with such modifications as that ε_0 represents the maximum electron energy in the band (in the degenerate state), rather than the value calculated for free electrons. In favourable cases, therefore, the width of the electron energy range can be approximately calculated from the susceptibility, as was pointed out by Peierls. This method must, of course, be applied with caution, for the free electron correspondence will break down if a discontinuity (a gap between bands) occurs within, or near the top of, the " occupied " energy range. Apart from this extended range of applicability of the free electron formulae, the formulae are of value in providing precise results for a limiting case which enable a check to be applied to the usually complicated formulae derived in attempts to take into account various types of perturbation.

2—Preliminary Formulae

Magnetic Moment—The magnetic moment is most conveniently calculated from the general formula

$$M = -(\partial F/\partial H)_{T,V} \tag{2.1}$$

where F is the free energy. With Fermi-Dirac statistics

$$F = NkT\eta + \Omega \tag{2.2}$$

$$\Omega = -kT\sum_s \log\{1 + \exp(\eta - \varepsilon_s/kT)\}. \tag{2.3}$$

Here η is written in place of the more usual $(-\alpha)$, as this simplifies the more frequently occurring subsequent formulae in removing (-1) raised to various powers. (It may be noted that η corresponds to ω/kT, or ζ/kT as used by Bloch,* and Bethe and Sommerfeld† respectively. The value of ηkT corresponds to the energy at which the proportion of occupied cells falls to half, and so in the completely degenerate state to ε_0.) The value of η in terms of N is obtained from

$$N = -\frac{1}{kT}\frac{\partial\Omega}{\partial\eta} = \frac{\partial}{\partial\eta}\sum_s \log\{1 + \exp(\eta - \varepsilon_s/kT)\}. \tag{2.4}$$

This corresponds to the well-known expression

$$N = \sum_s 1/\{\exp(\varepsilon_s/kT - \eta) + 1\}. \tag{2.4a}$$

From the above equation

$$\begin{aligned} M &= -\frac{\partial F}{\partial H} = -NkT\frac{\partial\eta}{\partial H} - \frac{\partial\Omega}{\partial\eta}\frac{\partial\eta}{\partial H} - \frac{\partial\Omega}{\partial H} \\ &= -(\partial\Omega/\partial H)_{\eta, T, V}\,. \end{aligned} \tag{2.5}$$

For the mean contıibution to the magnetic moment

$$\frac{M}{N} = kT\frac{\partial}{\partial H}\sum_s \log\{1 + \exp(\eta - \varepsilon_s/kT)\}/\sum_s 1/\{\exp(\varepsilon_s/kT - \eta) + 1\}. \tag{2.6}$$

Fermi-Dirac Integrals—The integrals occurring in applications of the Fermi-Dirac statistics are treated by Brillouin.‡ The results required in this paper are here given in a convenient form. The most frequently occurring integrals are of the form

$$F_n(\eta) = \int_0^\infty \frac{z^n\,dz}{e^{z-\eta} + 1}\,. \tag{2.7}$$

The following general relation holds:

$$\frac{\partial}{\partial\eta}\int_0^\infty \frac{z^n\,dz}{e^{z-\eta} + 1} = F'_n(\eta) = n\int_0^\infty \frac{z^{n-1}\,dz}{e^{z-\eta} + 1} = nF_{n-1}(\eta). \tag{2.8}$$

---

* 'Handb. der Radiologie,' vol. 6, p. 378 (1934).

† 'Handb. der Physik,' vol. 24 (2), p. 473 (1933).

‡ "Die Quantenstatistik," p. 503 (Berlin, 1931).

This can be shown by writing $nz^{n-1} = \frac{d}{dz} z^n$ and integrating by parts. Convenient expansions of (2.7) may be obtained for $\eta \gg 1$ (approach to completely degenerate state) and for $-\eta \gg 1$ (approach to classical statistics)

$\eta \gg 1$

$$F_n(\eta) = \int_0^\infty \frac{z^n\,dz}{e^{z-\eta}+1} = \frac{1}{n+1}\,\eta^{n+1} + \frac{\pi^2 n}{6}\,\eta^{n-1} \ldots \tag{2.9}$$

In particular (as the integral mainly required below)

$$\begin{aligned} F_{1/2}(\eta) = F(\eta) &= \int_0^\infty \frac{z^{1/2}\,dz}{e^{z-\eta}+1} = \tfrac{2}{3}\,\eta^{3/2} + \frac{\pi^2}{12}\,\eta^{-1/2} \ldots \\ &= \tfrac{2}{3}\,\eta^{3/2}\left(1 + \frac{\pi^2}{8\eta^2} \ldots\right). \end{aligned} \tag{2.10}$$

Other integrals can readily be obtained from (2.8). It may be noted that with increasing η, η tends to the limit

$$\eta = \varepsilon_0/kT, \tag{2.11}$$

$-\eta \gg 1$

$$F_n(\eta) = \{\Gamma(n+1)\}\left(e^\eta - \frac{e^{2\eta}}{2^{n+1}} + \frac{e^{3\eta}}{3^{n+1}} - \ldots\right). \tag{2.12}$$

In particular

$$F_{1/2}(\eta) = \frac{\sqrt{\pi}}{2}\left(e^\eta - \frac{e^{2\eta}}{2^{3/2}} + \frac{e^{3\eta}}{3^{3/2}} - \ldots\right). \tag{2.13}$$

Again, other integrals are obtainable from (2.8).
It may be noted that with decreasing η, e^η tends to the limit

$$e^\eta = \frac{4}{3\sqrt{\pi}}\left(\frac{\varepsilon_0}{kT}\right)^{3/2}. \tag{2.14}$$

Roughly, the ranges of validity of the expansions for $\eta \gg 1$ and $-\eta \gg 1$ correspond to $\varepsilon_0 \gg kT$ and $\varepsilon_0 \ll kT$.

For free electrons the value of ε_0 is given by

$$\varepsilon_0 = \frac{h^2}{2m}\left(\frac{3N}{8\pi V}\right)^{2/3} \tag{2.15}$$

$$\varepsilon_0/kT = 4{\cdot}20 \times 10^{-11}\,(N/V)^{2/3}/T. \tag{2.16}$$

For $N/V = 10^{22}$ — a rough average value for the quasi-free electron concentration in metals — $\varepsilon_0/kT = 1{\cdot}95 \times 10^4/T$.

Summation Formulae—In dealing with diamagnetism, it is convenient to convert certain sums into integrals. For this the Euler formula may be applied.

$$\sum_{n=a}^{n=b-1} f(n+\tfrac{1}{2}) = \int_a^b f(x)\,dx - \tfrac{1}{24}\,|f'(x)|_a^b. \tag{2.17}$$

An extension of this formula and its analogue for n in place of $(n+\frac{1}{2})$ have been given by Niessen.*

$$\sum_{n=a}^{n=b-1} f(n+\tfrac{1}{2}) = \int_a^b f(x)\,dx - \tfrac{1}{24}\,|f'(x) - \tfrac{7}{240}f'''(x)|_a^b. \tag{2.18}$$

$$\sum_{n=a}^{n=b} f(n) = \int_a^b f(x)\,dx + \tfrac{1}{2}f(a) + \tfrac{1}{2}f(b) + \tfrac{1}{12}\,|f'(x) - \tfrac{1}{60}f'''(x)|_a^b. \tag{2.19}$$

3—PARAMAGNETISM

The spin effect will first be considered apart from the diamagnetic effect. The energy states for the electron in the field are given by

$$\varepsilon = p^2/2m \pm \mu\mathrm{H}. \tag{3.1}$$

The number of states in the range dp for each direction of spin is $4\pi\mathrm{V}p^2\,dp/h^3$. The expression (2.3) for Ω becomes

$$\Omega = -(4\pi\mathrm{V}k\mathrm{T}/h^3)\int_0^\infty \Big[p^2 \log\Big\{1+\exp\Big(\eta - \frac{p^2}{2mk\mathrm{T}} - \frac{\mu\mathrm{H}}{k\mathrm{T}}\Big)\Big\} + p^2\log\Big\{1+\exp\Big(\eta - \frac{p^2}{2mk\mathrm{T}} + \frac{\mu\mathrm{H}}{k\mathrm{T}}\Big)\Big\}\Big]\,dp. \tag{3.2}$$

Let

$$\beta = \mu\mathrm{H}/k\mathrm{T} \tag{3.3}$$

$$z = p^2/2mk\mathrm{T} \tag{3.4}$$

$$\Omega = -(4\pi\mathrm{V}k\mathrm{T}/h^3)\,(mk\mathrm{T})^{3/2}\,2^{1/2}\int_0^\infty z^{1/2}\{\log(1+e^{\eta-z-\beta}) + \log(1+e^{\eta-z+\beta})\}\,dz. \tag{3.5}$$

This may be expressed as a power series in H in the form

$$\Omega = \Omega_0 + \frac{\beta^2}{2}\frac{\partial^2\Omega_0}{\partial\eta^2} + \frac{\beta^4}{24}\frac{\partial^4\Omega_0}{\partial\eta^4}\,\ldots, \tag{3.6}$$

where Ω_0 is the value of Ω for $\mathrm{H}=0$.

---

* 'Physica,' vol. 1, p. 783 (1934).

In terms of Ω_0 the values of M and N from (2.4) and (2.5) are given by

$$M = -\frac{\mu^2 H}{(kT)^2}\left\{\frac{\partial^2\Omega_0}{\partial\eta^2} + \frac{1}{6}\left(\frac{\mu H}{kT}\right)^2 \frac{\partial^4\Omega_0}{\partial\eta^4}\dots\right\} \tag{3.7}$$

$$N = -\frac{1}{kT}\left\{\frac{\partial\Omega_0}{\partial\eta} + \frac{1}{2}\left(\frac{\mu H}{kT}\right)^2 \frac{\partial^3\Omega_0}{\partial\eta^3}\dots\right\}. \tag{3.8}$$

In order to calculate M and N as power series in H or T a convenient expression for Ω_0 is required. Substituting the value (2.15) for ε_0 in (3.5) (with $\beta = 0$)

$$\Omega_0 = -\frac{3}{2} NkT\left(\frac{kT}{\varepsilon_0}\right)^{3/2}\int_0^\infty z^{1/2}\log\left(1+e^{\eta-z}\right)dz. \tag{3.9}$$

Integrating by parts

$$\begin{aligned}\Omega_0 &= -NkT\left(\frac{kT}{\varepsilon_0}\right)^{3/2}\int_0^\infty \frac{z^{3/2}\,dz}{e^{z-\eta}+1}\\ &= -NkT\left(\frac{kT}{\varepsilon_0}\right)^{3/2} F_{3/2}(\eta).\end{aligned} \tag{3.10}$$

It is convenient to use the following abbreviations:

$$F_{1/2}(\eta) = F \qquad \frac{\partial}{\partial\eta}F_{1/2}(\eta) = F', \text{ etc.} \tag{3.11}$$

Using (3.6), (3.11), and (2.8)

$$\Omega = -NkT\left(\frac{kT}{\varepsilon_0}\right)^{3/2}\left[F_{3/2}(\eta) + \frac{3\beta^2}{4}F'(\eta) + \frac{\beta^4}{16}F'''(\eta)\dots\right]. \tag{3.12}$$

Using (2.4), (2.5), and (3.12), or (3.7), (3.8), and (3.10), the expressions for M and N to the second approximation become

$$M = \frac{3}{2}\frac{N\mu^2 H}{kT}\left(\frac{kT}{\varepsilon_0}\right)^{3/2}\left\{F' + \frac{\beta^2}{6}F'''\dots\right\} \tag{3.13}$$

$$N = \frac{3}{2}N\left(\frac{kT}{\varepsilon_0}\right)^{3/2}\left\{F + \frac{\beta^2}{2}F''\dots\right\}. \tag{3.14}$$

To the first power of H the simple result is thus obtained

$$\frac{M}{N} = \frac{\mu^2 H}{kT}\frac{F'}{F}. \tag{3.15}$$

The appropriate expansions for low and high temperatures will now be derived.

(i) $\underline{\varepsilon_0 \gg kT}$—From (2.9) the following asymptotic expansions are obtained for F and its derivatives.

$\eta \gg 1$

$$\left.\begin{aligned} F &= \tfrac{2}{3}\eta^{3/2}\left(1 + \frac{\pi^2}{8\eta^2}\cdots\right) \\ F' &= \eta^{1/2}\left(1 - \frac{\pi^2}{24\eta^2}\cdots\right) \\ F'' &= \tfrac{1}{2}\eta^{-1/2}\left(1 + \frac{\pi^2}{8\eta^2}\cdots\right) \\ F''' &= -\tfrac{1}{4}\eta^{-3/2}\left(1 - \frac{5\pi^2}{8\eta^2}\cdots\right) \end{aligned}\right\} . \tag{3.16}$$

In the terms involving β^2 in (3.13) and (3.14) only the first term in the above expansions need be retained.

$$M = \tfrac{3}{2}\frac{N\mu^2H}{kT}\left(\frac{kT}{\varepsilon_0}\right)^{3/2}\eta^{1/2}\left\{1 - \frac{\pi^2}{24\eta^2} - \frac{\beta^2}{24\eta^2}\cdots\right\}, \tag{3.17}$$

$$N = \tfrac{3}{2}N\left(\frac{kT}{\varepsilon_0}\right)^{3/2}\tfrac{2}{3}\eta^{3/2}\left\{1 + \frac{\pi^2}{8\eta^2} + \frac{3\beta^2}{8\eta^2}\cdots\right\}. \tag{3.18}$$

From (3.18)

$$\eta = \left(\frac{\varepsilon_0}{kT}\right)\left\{1 + \frac{\pi^2}{8\eta^2} + \frac{3\beta^2}{8\eta^2}\cdots\right\}^{-2/3}, \tag{3.19}$$

giving

$$\begin{aligned} M &= \tfrac{3}{2}\frac{N\mu^2H}{\varepsilon_0}\left\{1 - \frac{\pi^2}{12\eta^2} - \frac{\beta^2}{6\eta^2}\cdots\right\} \\ &\doteqdot \tfrac{3}{2}\frac{N\mu^2H}{\varepsilon_0}\left\{1 - \frac{\pi^2}{12}\left(\frac{kT}{\varepsilon_0}\right)^2 - \tfrac{1}{6}\left(\frac{\mu H}{\varepsilon_0}\right)^2\right\}. \end{aligned} \tag{3.20}$$

This result is in disagreement with Niessen's result* for the dependence on T, and agrees with that of Bloch.† It agrees with Niessen's result‡ for the dependence on H.

(ii) $\underline{\varepsilon_0 \ll kT}$—From (2.13) the following expansions are obtained:

$-\eta \gg 1$

$$\left.\begin{aligned} F &= \frac{\sqrt{\pi}}{2}e^{\eta}\left(1 - \frac{e^{\eta}}{2^{3/2}}\cdots\right) \\ F' &= \frac{\sqrt{\pi}}{2}e^{\eta}\left(1 - \frac{e^{\eta}}{2^{1/2}}\cdots\right) \\ F'' &= \frac{\sqrt{\pi}}{2}e^{\eta}\left(1 - 2^{1/2}e^{\eta}\right) \\ F''' &= \frac{\sqrt{\pi}}{2}e^{\eta}\left(1 - 2^{3/2}e^{\eta}\ldots\right) \end{aligned}\right\} . \tag{3.21}$$

---

* 'Physica,' vol. 1, p. 979 (1934).

† 'Z. Physik,' vol. 53, p. 216 (1929).

‡ 'Physica,' vol. 1, p. 783 (1934).

Again, only the first term in the above expansions need be retained in the terms in (3.13) and (3.14) involving β^2.

$$M = \tfrac{3}{2}\frac{N\mu^2H}{kT}\left(\frac{kT}{\varepsilon_0}\right)^{3/2}\frac{\sqrt{\pi}}{2}e^{\eta}\left\{1-\frac{e^{\eta}}{2^{1/2}}+\frac{\beta^2}{6}\ldots\right\}. \tag{3.22}$$

$$N = \tfrac{3}{2}N\left(\frac{kT}{\varepsilon_0}\right)^{3/2}\frac{\sqrt{\pi}}{2}e^{\eta}\left\{1-\frac{e^{\eta}}{2^{3/2}}+\frac{\beta^2}{2}\ldots\right\}. \tag{3.23}$$

From (3.23)

$$e^{\eta} = \frac{4}{3\sqrt{\pi}}\left(\frac{\varepsilon_0}{kT}\right)^{3/2}\left\{1-\frac{e^{\eta}}{2^{3/2}}+\frac{\beta^2}{2}\ldots\right\}^{-1}, \tag{3.24}$$

giving

$$M = \frac{N\mu^2H}{kT}\left\{1-\frac{e^{\eta}}{2^{3/2}}-\frac{\beta^2}{3}\ldots\right\} \tag{3.25}$$

$$\doteqdot \frac{N\mu^2H}{kT}\left\{1-\tfrac{1}{3}\left(\frac{2}{\pi}\right)^{1/2}\left(\frac{\varepsilon_0}{kT}\right)^{3/2}-\tfrac{1}{3}\left(\frac{\mu H}{kT}\right)^2\right\}. \tag{3.26}$$

This expression shows that the susceptibility falls below the value given by the classical result as the temperature falls. The term involving H^3 is the second term in the expansion of tanh $(\mu H/kT)$. The exact classical result is readily obtained directly from (3.5).

For $e^{\eta-z\pm\beta} \ll 1$

$$\log\left(1+e^{\eta-z\pm\beta}\right) \doteqdot e^{\eta-z\pm\beta}, \tag{3.27}$$

giving

$$\Omega = -(4\pi VkT/h^3)(mkT)^{3/2}\,2^{1/2}e^{\eta}\left(e^{-\beta}+e^{+\beta}\right)\int_0^{\infty} z^{1/2}e^{-z}\,dz$$

$$= -(VkT/h^3)(2\pi mkT)^{3/2}e^{\eta}\left(e^{-\beta}+e^{+\beta}\right). \tag{3.28}$$

From (2.4) and (2.5)

$$\frac{M}{N} = kT\frac{(\partial\Omega/\partial H)}{(\partial\Omega/\partial\eta)}$$

$$= \frac{kT\mu}{kT}\left(\frac{-e^{-\beta}+e^{\beta}}{e^{-\beta}+e^{\beta}}\right)$$

$$= \mu\tanh\beta \doteqdot \frac{\mu^2H}{kT}\left\{1-\tfrac{1}{3}\left(\frac{\mu H}{kT}\right)^2\right\}. \tag{3.29}$$

This method corresponds to the usual one in which the moment is derived from the classical partition function.

$$M = NkT\frac{\partial}{\partial H}\left(\log\Sigma\, e^{-\varepsilon_i/kT}\right). \tag{3.30}$$

4—DIAMAGNETISM

It is unnecessary to enter into detail as to the derivation of the characteristic energy values for a free electron (without spin) in a magnetic field, as this has been discussed at length in several of the papers referred to above. There are various possible presentations of the treatment, superficially different, but equivalent in leading to the same expression for Ω, or for the classical partition function. The following outline, based on the presentation by Sommerfeld and Bethe,* will be sufficient.

The wave equation, as modified for the presence of a magnetic field (the scalar potential being zero), takes the form

$$E\psi - \frac{1}{2m}\sum_k\left(\frac{h}{2\pi i}\frac{\partial}{\partial q_k} - \frac{e}{c}A_k\right)^2\psi = 0. \tag{4.1}$$

A field H along the z axis may be derived from a vector potential of components

$$A_x = -Hy, \quad A_y = A_z = 0. \tag{4.2}$$

The wave equation then becomes

$$\nabla^2\psi + iH\frac{4\pi e}{hc}\frac{\partial\psi}{\partial x} + \frac{8\pi^2 m}{h^2}\left(E - \frac{e^2}{2mc^2}H^2y^2\right)\psi = 0. \tag{4.3}$$

This equation is solved by

$$\psi = \phi(y)\exp\left\{\frac{2\pi i}{h}(-p_x x + p_z z)\right\}, \tag{4.4}$$

giving as the equation for $\phi(y)$

$$\frac{\partial^2\phi}{\partial y^2} + \frac{8\pi^2}{h^2}\left\{\left(E - \frac{p_z^2}{2m}\right) - \frac{1}{2m}\left(\frac{eH}{c}y - p_x\right)^2\right\}\phi = 0. \tag{4.5}$$

This corresponds to the equation for a harmonic oscillator with equilibrium position at

$$y = cp_x/eH \tag{4.6}$$

and frequency

$$\nu_0 = eH/(2\pi mc). \tag{4.7}$$

The proper values of the original equation are therefore

$$E = \frac{p_z^2}{2m} + h\nu_0(n + \tfrac{1}{2})$$

$$= \frac{p_z^2}{2m} + (2n + 1)H\mu. \tag{4.8}$$

The number of states for a given value of n, and for a range dp_z are next required. Let A, B, C be the linear dimensions of the metal in the x, y, z directions, $\nu(n)$ the number of states per unit range.

---

* 'Handb. der Physik,' vol. 24 (2), p. 473 (1933).

$$\begin{aligned}\nu(n)\,dp_z &= (\mathrm{AC}/h^2)\,dp_z \int dp_x \\ &= \frac{e\mathrm{H}}{ch^2}\,\mathrm{AC}\,dp_z \int_0^{\mathrm{B}} dy \\ &= \mathrm{V}\frac{e\mathrm{H}}{ch^2}\,dp_z. \qquad (4.9)\end{aligned}$$

Use is here made of the relation (4.6), and the condition that y must be less than B if the calculated characteristic values are to remain valid. The error involved, as shown by Landau, becomes negligible if B is sufficiently large; practically, the necessary conditions will always be fulfilled.

For electrons, the above value of $\nu(n)$ must be multiplied by the statistical weight factor 2. The subscript z may be omitted. The function Ω then becomes

$$\Omega = -k\mathrm{TV}\frac{2e\mathrm{H}}{h^2c}\sum_{n=0}^{\infty}\int_{-\infty}^{+\infty}\log\left[1+\exp\left\{\eta-\frac{p^2}{2mk\mathrm{T}}-(2n+1)\beta\right\}\right]dp. \qquad (4.10)$$

Making the substitution $p^2 = 2mk\mathrm{T}z$, and proceeding as in connexion with (3.9),

$$\Omega = -\tfrac{3}{2}\mathrm{N}\mu\mathrm{H}\left(\frac{k\mathrm{T}}{\varepsilon_0}\right)^{3/2}\sum_{n=0}^{\infty}\int_0^{\infty} z^{-1/2}\log\{1+e^{\eta-z-(2n+1)\beta}\}\,dz. \qquad (4.11)$$

Integrating by parts

$$\begin{aligned}\Omega &= -3\mathrm{N}\mu\mathrm{H}\left(\frac{k\mathrm{T}}{\varepsilon_0}\right)^{3/2}\sum_{n=0}^{\infty}\int_0^{\infty}\frac{z^{1/2}\,dz}{\exp\{z+(2n+1)\beta-\eta\}+1} \\ &= -3\mathrm{N}\mu\mathrm{H}\left(\frac{k\mathrm{T}}{\varepsilon_0}\right)^{3/2}\sum_{n=0}^{\infty}\mathrm{F}_{1/2}\{\eta-(2n+1)\beta\}. \qquad (4.12)\end{aligned}$$

The sum may now be evaluated by the use of the formula (2.18)

$$\begin{aligned}\sum_{n=0}^{\infty}\mathrm{F}_{1/2}\{\eta-(n+\tfrac{1}{2})2\beta\} &= \int_0^{\infty}\mathrm{F}_{1/2}(\eta-2\beta x)\,dx \\ &\quad -\tfrac{1}{24}\left|\frac{\partial}{\partial x}\mathrm{F}(\eta-2\beta x)-\tfrac{7}{240}\frac{\partial^3}{\partial x^3}\mathrm{F}(\eta-2\beta x)\right|_0^{\infty}. \qquad (4.13)\end{aligned}$$

To evaluate the integral, let

$$\eta - 2\beta x = \eta'$$

$$\int_0^\infty F_{1/2}(\eta - 2\beta x) = -\frac{1}{2\beta}\int_\eta^{-\infty} F_{1/2}(\eta')\,d\eta'$$

$$= \frac{1}{3\beta} F_{3/2}(\eta).$$

The two derivatives also vanish at the upper limit.

$$\left|\frac{\partial}{\partial x} F(\eta - 2\beta x)\right|_0^\infty = 2\beta F'(\eta)$$

$$\left|\frac{\partial^3}{\partial x^3} F(\eta - 2\beta x)\right|_0^\infty = 8\beta^3 F'''(\eta).$$

Substituting these values, the expression finally obtained for Ω is

$$\Omega = -NkT\left(\frac{kT}{\varepsilon_0}\right)^{3/2}\left[F_{3/2}(\eta) - \tfrac{1}{4}\beta^2 F'(\eta) + \tfrac{7}{240}\beta^4 F'''(\eta)\ldots\right]. \quad (4.14)$$

This expression for Ω, taking into account the diamagnetic effect alone, may be compared with (3.12), taking into account the spin effect alone.

To the second approximation the values of M and N are given by

$$M = -\frac{\partial\Omega}{\partial H} = -\tfrac{1}{2}\frac{N\mu^2 H}{kT}\left(\frac{kT}{\varepsilon_0}\right)^{3/2}\left\{F' - \frac{7\beta^2}{30}F'''\right\}, \quad (4.15)$$

$$N = -\frac{1}{kT}\frac{\partial\Omega}{\partial\eta} = \tfrac{3}{2}N\left(\frac{kT}{\varepsilon_0}\right)^{3/2}\left\{F - \frac{\beta^2}{6}F''\right\}. \quad (4.16)$$

The appropriate expansions for low and high temperatures are now readily obtained.

(i) $\underline{\varepsilon_0 \gg kT}$—Using (3.16)

$$M = -\tfrac{1}{2}\frac{N\mu^2 H}{kT}\left(\frac{kT}{\varepsilon_0}\right)^{3/2}\eta^{1/2}\left\{1 - \frac{\pi^2}{24\eta^2} + \frac{7\beta^2}{120\eta^2}\right\}, \quad (4.17)$$

$$N = N\left(\frac{kT}{\varepsilon_0}\right)^{3/2}\eta^{3/2}\left\{1 + \frac{\pi^2}{8\eta^2} - \frac{\beta^2}{8\eta^2}\right\}. \quad (4.18)$$

From (4.18)

$$\eta = \left(\frac{\varepsilon_0}{kT}\right)\left\{1 + \frac{\pi^2}{8\eta^2} - \frac{\beta^2}{8\eta^2}\right\}^{-2/3}, \quad (4.19)$$

giving

$$M = -\tfrac{1}{2}\frac{N\mu^2 H}{\varepsilon_0}\left\{1 - \frac{\pi^2}{12}\left(\frac{kT}{\varepsilon_0}\right)^2 + \tfrac{1}{10}\left(\frac{\mu H}{\varepsilon_0}\right)^2\right\}. \quad (4.20)$$

This result again agrees with Niessen's for the dependence on H, but disagrees for the dependence on T.

(ii) $\underline{\varepsilon_0 \ll kT}$—Using (3.21)

$$M = -\tfrac{1}{2}\frac{N\mu^2 H}{kT}\left(\frac{kT}{\varepsilon_0}\right)^{3/2}\frac{\sqrt{\pi}}{2}e^{\eta}\left\{1 - \frac{e^{\eta}}{2^{1/2}} - \frac{7\beta^2}{30}\right\}, \tag{3.21}$$

$$N = \tfrac{3}{2}N\left(\frac{kT}{\varepsilon_0}\right)^{3/2}\frac{\sqrt{\pi}}{2}e^{\eta}\left\{1 - \frac{e^{\eta}}{2^{3/2}} - \frac{\beta^2}{6}\right\}. \tag{3.22}$$

From (3.22)

$$e^{\eta} = \frac{4}{3\sqrt{\pi}}\left(\frac{\varepsilon_0}{kT}\right)^{3/2}\left\{1 - \frac{e^{\eta}}{2^{3/2}} - \frac{\beta^2}{6}\right\}^{-1}, \tag{3.23}$$

giving

$$M = -\tfrac{1}{3}\frac{N\mu^2 H}{kT}\left\{1 - \tfrac{1}{3}\left(\frac{2}{\pi}\right)^{1/2}\left(\frac{\varepsilon_0}{kT}\right)^{3/2} - \tfrac{1}{15}\left(\frac{\mu H}{kT}\right)^2\right\}. \tag{3.24}$$

The result corresponding to classical statistics may be obtained from (4.10) in a manner similar to that used in deriving (3.29). It is unnecessary to give details. The result is

$$\begin{aligned} M &= -N\mu\left\{\coth\left(\frac{\mu H}{kT}\right) - \left(\frac{kT}{\mu H}\right)\right\} \\ &\doteqdot -\tfrac{1}{3}\frac{N\mu^2 H}{kT}\left\{1 - \tfrac{1}{15}\left(\frac{\mu H}{kT}\right)^2\right\}, \end{aligned} \tag{3.25}$$

in agreement with (3.24) for the term in H^3.

5—RESULTANT EFFECT

The resultant effect, when the spin and diamagnetic effect are considered together, is not necessarily the simple sum of the two effects considered separately. The possible magnetic energies of the electrons are given by

$$\varepsilon = (2n + 1)\,\mu H \pm \mu H \tag{5.1}$$

and Ω, in the form corresponding to (4.12), becomes

$$\Omega = -\tfrac{3}{2}N\mu H\left(\frac{kT}{\varepsilon_0}\right)^{3/2}\sum_{n=0}^{\infty}[F_{1/2}\{\eta - (n+1)\,2\beta\} + F_{1/2}\{\eta - 2n\beta\}], \tag{5.2}$$

$$\sum_{n=0}^{\infty} = 2\,[\sum_{n=0}^{\infty}\{F_{1/2}(\eta - 2n\beta)\} - \tfrac{1}{2}F_{1/2}(\eta)]. \tag{5·3}$$

The sum may be evaluated, using Niessen's formula (2.19), by a method similar to that used in proceeding from (4.12) to (4.14). The final result obtained for Ω is

$$\Omega = -\mathrm{N}k\mathrm{T}\left(\frac{k\mathrm{T}}{\varepsilon_0}\right)^{3/2}\left[\mathrm{F}_{3/2}(\eta) + \tfrac{1}{2}\beta^2\mathrm{F}'(\eta) - \frac{\beta^4}{30}\mathrm{F}'''(\eta)\right], \tag{5.4}$$

giving

$$\mathrm{M} = \frac{\mathrm{N}\mu^2\mathrm{H}}{k\mathrm{T}}\left(\frac{k\mathrm{T}}{\varepsilon_0}\right)^{3/2}\left\{\mathrm{F}' - \frac{2\beta^2}{15}\mathrm{F}'''\right\}, \tag{5.5}$$

$$\mathrm{N} = \tfrac{3}{2}\mathrm{N}\left(\frac{k\mathrm{T}}{\varepsilon_0}\right)^{3/2}\left\{\mathrm{F} + \frac{\beta^2}{3}\mathrm{F}''\right\}. \tag{5.6}$$

(i) $\underline{\varepsilon_0 \gg k\mathrm{T}}$

$$\mathrm{M} = \frac{\mathrm{N}\mu^2\mathrm{H}}{k\mathrm{T}}\left(\frac{k\mathrm{T}}{\varepsilon_0}\right)^{3/2}\eta^{1/2}\left\{1 - \frac{\pi^2}{24\eta^2} + \frac{\beta^2}{30\eta^2}\right\} \tag{5.7}$$

$$\mathrm{N} = \mathrm{N}\left(\frac{k\mathrm{T}}{\varepsilon_0}\right)^{3/2}\eta^{3/2}\left\{1 + \frac{\pi^2}{8\eta^2} + \frac{\beta^2}{4\eta^2}\right\} \tag{5.8}$$

$$\eta = \left(\frac{\varepsilon_0}{k\mathrm{T}}\right)\left(1 + \frac{\pi^2}{8\eta^2} + \frac{\beta^2}{4\eta^2}\right)^{-2/3} \tag{5.9}$$

$$\mathrm{M} = \frac{\mathrm{N}\mu^2\mathrm{H}}{\varepsilon_0}\left\{1 - \frac{\pi^2}{12}\left(\frac{k\mathrm{T}}{\varepsilon_0}\right)^2 - \tfrac{1}{20}\left(\frac{\mu\mathrm{H}^2}{\varepsilon_0}\right)^2\right\}. \tag{5.10}$$

The temperature factor is the same as that found for the dia- and paramagnetic effects alone in (3.20) and (4.20), again with a disagreement in sign as compared with Niessen's result*; for the dependence on H there is agreement with Niessen.† To the first power in H and the second power in T, the resultant effect is the sum of the two effects considered separately; but this does not hold for the term in H^3, presumably owing to the manner in which η depends on H.

(ii) $\underline{\varepsilon_0 \ll k\mathrm{T}}$

$$\mathrm{M} = \frac{\mathrm{N}\mu^2\mathrm{H}}{k\mathrm{T}}\left(\frac{k\mathrm{T}}{\varepsilon_0}\right)^{3/2}\frac{\sqrt{\pi}}{2}e^{\eta}\left\{1 - \frac{e^{\eta}}{2^{1/2}} - \frac{2\beta^2}{15}\right\} \tag{5.11}$$

$$\mathrm{N} = \tfrac{3}{2}\mathrm{N}\left(\frac{k\mathrm{T}}{\varepsilon_0}\right)^{3/2}\frac{\sqrt{\pi}}{2}e^{\eta}\left\{1 - \frac{e^{\eta}}{2^{3/2}} + \frac{\beta^2}{3}\right\} \tag{5.12}$$

$$e^{\eta} = \frac{4}{3\sqrt{\pi}}\left(\frac{\varepsilon_0}{k\mathrm{T}}\right)^{3/2}\left\{1 - \frac{e^{\eta}}{2^{3/2}} + \frac{\beta^2}{3}\right\}^{-1} \tag{5.13}$$

$$\mathrm{M} = \tfrac{2}{3}\frac{\mathrm{N}\mu^2\mathrm{H}}{k\mathrm{T}}\left\{1 - \tfrac{1}{3}\left(\frac{2}{\pi}\right)^{1/2}\left(\frac{\varepsilon_0}{k\mathrm{T}}\right)^{3/2} - \tfrac{7}{15}\left(\frac{\mu\mathrm{H}}{k\mathrm{T}}\right)^2\right\}. \tag{5.14}$$

---

* 'Physica,' vol. 1, p. 979 (1934).

† *Ibid.*, vol. 1, p. 783 (1934).

The term in H^3 corresponds to that obtained with classical statistics, namely,

$$M = N\mu\left\{\tanh\beta - \left(\coth\beta - \frac{1}{\beta}\right)\right\}$$
$$\doteqdot \tfrac{2}{3}\frac{N\mu^2H}{kT}\left\{1 - \tfrac{7}{15}\left(\frac{\mu H}{kT}\right)^2\right\}. \qquad (5.15)$$

6—Collected Results. Numerical Expressions

For convenient reference, the results for the magnetic moment are collected together. Let M_P be the moment due to the paramagnetic effect, M_D to the diamagnetic effect (each considered alone), and M_R to the resultant effect.

(i) $\varepsilon_0 \gg kT$

$$M_P = \tfrac{3}{2}\frac{N\mu^2H}{\varepsilon_0}\left\{1 - \frac{\pi^2}{12}\left(\frac{kT}{\varepsilon_0}\right)^2 - \tfrac{1}{6}\left(\frac{\mu H}{\varepsilon_0}\right)^2\right\} \qquad (6.1)$$

$$M_D = -\tfrac{1}{2}\frac{N\mu^2H}{\varepsilon_0}\left\{1 - \frac{\pi^2}{12}\left(\frac{kT}{\varepsilon_0}\right)^2 + \tfrac{1}{10}\left(\frac{\mu H}{\varepsilon_0}\right)^2\right\} \qquad (6.2)$$

$$M_R = \frac{N\mu^2H}{\varepsilon_0}\left\{1 - \frac{\pi^2}{12}\left(\frac{kT}{\varepsilon_0}\right)^2 - \tfrac{1}{20}\left(\frac{\mu H}{\varepsilon_0}\right)^2\right\}. \qquad (6.3)$$

(ii) $\varepsilon_0 \ll kT$

$$M_P = \frac{N\mu^2H}{kT}\left\{1 - \tfrac{1}{3}\left(\frac{2}{\pi}\right)^{1/2}\left(\frac{\varepsilon_0}{kT}\right)^{3/2} - \tfrac{1}{3}\left(\frac{\mu H}{kT}\right)^2\right\} \qquad (6.4)$$

$$M_D = -\tfrac{1}{3}\frac{N\mu^2H}{kT}\left\{1 - \tfrac{1}{3}\left(\frac{2}{\pi}\right)^{1/2}\left(\frac{\varepsilon_0}{kT}\right)^{3/2} - \tfrac{1}{15}\left(\frac{\mu H}{kT}\right)^2\right\} \qquad (6.5)$$

$$M_R = \tfrac{2}{3}\frac{N\mu^2H}{kT}\left\{1 - \tfrac{1}{3}\left(\frac{2}{\pi}\right)^{1/2}\left(\frac{\varepsilon_0}{kT}\right)^{3/2} - \tfrac{7}{15}\left(\frac{\mu H}{kT}\right)^2\right\}. \qquad (6.6)$$

In the above expressions, for free electrons

$$\varepsilon_0 = \frac{h^2}{2m}\left(\frac{3N}{8\pi V}\right)^{2/3}. \qquad (6.7)$$

With classical statistics ($\varepsilon_0 = 0$)

$$M_R = M_P + M_D = N\mu\left[\tanh\left(\frac{\mu H}{kT}\right) - \left\{\coth\left(\frac{\mu H}{kT}\right) - \left(\frac{kT}{\mu H}\right)\right\}\right]. \qquad (6.8)$$

It may be noticed that the dependence on T of the resultant effect is the same as that of the effects considered separately, both for $\varepsilon_0 \gg kT$ and $\varepsilon_0 \ll kT$. To the approximation considered, the resultant effect is the sum of the separate effects except for the term involving H^3 for $\varepsilon_0 \gg kT$.

To enable the order of magnitude of the free electron effect to be more readily appreciated, a different method of expressing the results may be adopted, which will be applied to the results for the resultant moment, considering only the cases in which $(\mu H/kT)$ or $(\mu H/\varepsilon_0)$ is much less than unity.

Let $(\chi_A)_e$ be the gram atomic susceptibility due to the quasi-free electrons, q the number of these per atom, V_0 the value of ε_0 expressed in equivalent volts, N_A the number of atoms, so that $N = N_A q$.

$$\left.\begin{aligned} (\chi_A)_e &= (M_R/NH) \times q \times 6{\cdot}06 \times 10^{23} \\ 1 \text{ electron volt} &= 1{\cdot}59_1 \times 10^{-12} \text{ erg} \\ kT &= 8{\cdot}62_1 \times 10^{-5}\, T \text{ volt} \\ &= 1{\cdot}37_1 \times 10^{-16}\, T \text{ erg} \\ \varepsilon_0 \,(\text{erg}) &= 1{\cdot}59_1 \times 10^{-12}\, V_0 \,(\text{volt}) \\ \mu &= 9{\cdot}17_4 \times 10^{-21} \text{ erg gauss}^{-1} \end{aligned}\right\} . \tag{6.8}$$

$\underline{\varepsilon_0 \gg kT}$ $\qquad T \ll 1{\cdot}157 \times 10^4\, V_0$

$$\begin{aligned} (\chi_A)_e \times 10^6 &= \frac{6{\cdot}06 \times 10^{23} \mu^2 q}{\varepsilon_0} \left\{1 - \frac{\pi^2}{12}\left(\frac{kT}{\varepsilon_0}\right)^2\right\} \times 10^6 \\ &= 32{\cdot}1\,(q/V_0)\,\{1 - 6{\cdot}11 \times 10^{-9}\,(T/V_0)^2\}. \end{aligned} \tag{6.9}$$

$\underline{\varepsilon_0 \ll kT}$ $\qquad T \gg 1{\cdot}157 \times 10^4\, V_0$

$$(\chi_A)_e \times 10^6 = q \times \frac{248}{T} \times 10^3\,\{1 - 3{\cdot}32_3 \times 10^5\,(V_0/T)^{3/2}\}. \tag{6.10}$$

Finally, for free electrons

$$V_0 = 3{\cdot}618 \times 10^{-15}\,(qn_v)^{2/3}, \tag{6.11}$$

where n_v is the number of atoms per unit volume. (As indicative of orders of magnitude, the free electron V_0 for sodium is $3{\cdot}15$ volts, for silver $5{\cdot}5$ volts.)

The form (6.11) may be taken as an approximation for electrons in an unfilled band, provided that ε_0 is considerably less than the total band width ; in general, the numerical factor will be smaller.

It may be noted that the expressions derived for the magnetic moment may all be put in a form which brings out the fact that the moment acquired in a field depends essentially on the energy distribution of the electrons only in the neighbourhood of ηkT (that is, the value of ε at which the proportion of cells occupied is one-half). From a purely theoretical point of view this is preferable, owing to the increase in generality ; but the increase in generality is formal, for it is not possible to make calculations of the moment unless the energy distribution in the neighbourhood of ηkT is known. The formulae as given in this section apply strictly to free electrons ; but they have a wider applicability as long as the relation between electron energy and momentum is the same (apart from numerical factors) as for free electrons.

A graphical representation of the results is given in fig. 1.

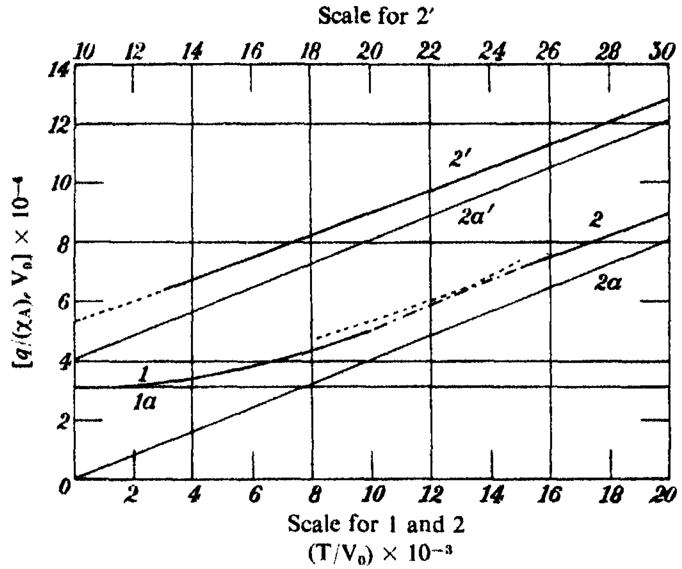


FIG. 1—Variation with temperature of electron susceptibility.

$(\chi_A)_e$ = gm atomic susceptibility due to electrons.

q = number of quasi-free electrons per atom.

V_0 = maximum electron energy in completely degenerate state in equivalent volts.

The full curves 1 and 2 are drawn from the formulae for $\varepsilon_0 \gg kT$ and $\varepsilon_0 \ll kT$. 1*a* and 2*a* are the corresponding limiting curves. $\varepsilon_0 = kT$ for $T/V_0 = 1 \cdot 157 \times 10^4$. The dotted curves are continuations of the full curves beyond the range of validity of the formulae. The dot dash curve is drawn to connect the two full curves in a continuous manner.

7—CRITICISM OF NIESSEN'S TREATMENT THERMODYNAMIC FUNCTIONS

The question arises as to why the results obtained here for the temperature variation of susceptibility for $\varepsilon_0 \gg kT$ differ from those of Niessen.*

---

* 'Physica,' vol. 1, p. 979 (1934).

Niessen derives expressions for the total energy of the electron gas, and from them he derives the magnetic moment by applying the formula

$$\mathrm{M} = -(\partial\mathrm{E}/\partial\mathrm{H}). \tag{7.1}$$

Now this formula holds for constant entropy, the thermodynamic equations (omitting pressure volume terms) being

$$d\mathrm{E} = \mathrm{T}d\mathrm{S} - \mathrm{M}d\mathrm{H} \tag{7.2}$$

$$d\mathrm{F} = -\mathrm{S}d\mathrm{T} - \mathrm{M}d\mathrm{H}. \tag{7.3}$$

Niessen, however, in effect carries out the partial differentiation with T constant. Denoting the result he obtains by M′

$$\mathrm{M}' = -(\partial\mathrm{E}/\partial\mathrm{H})_{\mathrm{T}} = \mathrm{M} - \mathrm{T}\,(\partial\mathrm{S}/\partial\mathrm{H})_{\mathrm{T}}. \tag{7.4}$$

The thermodynamic relations have been discussed in a previous paper.* Applying the relation derived from (7.3)

$$(\partial\mathrm{S}/\partial\mathrm{H})_{\mathrm{T}} = (\partial\mathrm{M}/\partial\mathrm{T})_{\mathrm{H}} \tag{7.5}$$

$$\mathrm{M}' = \mathrm{M} - \mathrm{T}\,(\partial\mathrm{M}/\partial\mathrm{T})_{\mathrm{H}}. \tag{7.6}$$

Using the value found above for M_{R} (6.3)

$$\begin{aligned}\mathrm{M}'_{\mathrm{R}} &= \frac{\mathrm{N}\mu^2\mathrm{H}}{\varepsilon_0}\left[\left\{1 - \frac{\pi^2}{12}\left(\frac{k\mathrm{T}}{\varepsilon_0}\right)^2\right\} - \mathrm{T}\,\frac{\partial}{\partial\mathrm{T}}\left\{1 - \frac{\pi^2}{12}\left(\frac{k\mathrm{T}}{\varepsilon_0}\right)^2\right\}\right] \\ &= \frac{\mathrm{N}\mu^2\mathrm{H}}{\varepsilon_0}\left\{1 + \frac{\pi^2}{12}\left(\frac{k\mathrm{T}}{\varepsilon_0}\right)^2\right\}.\end{aligned} \tag{7.7}$$

The erroneous result obtained by Niessen is thus not due to errors in the development of the treatment, but to an incorrect application of (7.1). (The incorrect result obtained in treating the paramagnetism separately (*loc. cit.*, p. 985),† by a method which is valid, appears to be due to an algebraical slip.)

The expression for the free energy F may be obtained from (2.2) and the expressions derived for η and Ω. Considering both the dia- and paramagnetic effects, the result derived for $\varepsilon_0 \gg k\mathrm{T}$ as far as second powers in H and T is

$$\mathrm{F} = \tfrac{3}{5}\,\mathrm{N}\varepsilon_0\left[1 - \frac{5\pi^2}{12}\left(\frac{k\mathrm{T}}{\varepsilon_0}\right)^2 - \tfrac{5}{6}\left(\frac{\mu\mathrm{H}}{\varepsilon_0}\right)^2\left\{1 - \frac{\pi^2}{12}\left(\frac{k\mathrm{T}}{\varepsilon_0}\right)^2\right\}\right]. \tag{7.8}$$

---

* Stoner, 'Phil. Mag.,' vol. 19, p. 565 (1935).

† 'Physica,' vol. 1, p. 979 (1934).

From this the moment may be derived from (2.1). For the total energy

$$\begin{aligned} E &= F - T\,(\partial F/\partial T) \\ &= \tfrac{3}{5}\,N\varepsilon_0\left[1 + \frac{5\pi^2}{12}\left(\frac{kT}{\varepsilon_0}\right)^2 - \tfrac{5}{6}\left(\frac{\mu H}{\varepsilon_0}\right)^2\left\{1 + \frac{\pi^2}{12}\left(\frac{kT}{\varepsilon_0}\right)^2\right\}\right]. \end{aligned} \qquad (7.9)$$

*This expression for E is in agreement with Niessen's.**

8—The Experimental Results

It is unnecessary to consider the experimental results on the susceptibility of metals here, as they have been discussed in some detail in "Magnetism and Matter."† It is there shown that the susceptibilities of a large number of metals may be accounted for as arising from a paramagnetic contribution which varies little with temperature, and a diamagnetic contribution due to the electrons in completed groups in the atoms. (This simple method of treating the diamagnetism does not, of course, apply to such metals as bismuth.) The values derived for V_0 by using the first term of (6.9) are generally considerably less than the values for free electrons ; but the band widths indicated are of an order of magnitude which fits in satisfactorily with other evidence.

There is one point which may pertinently be mentioned here. For the alkali metals the conditions approach most closely those for which the free electron calculations should apply. From the results of Sucksmith,‡ however, the susceptibility increases slightly with temperature, whereas (6.9) indicates a decrease. Actually, the decrease, as given by the formula, would be inappreciable. It is suggested that the increase for these metals is mainly due to the change in the value of V_0 (or ε_0) associated with thermal expansion. From (6.11)

$$V_0 \propto n_v^{2/3}. \qquad (8.1)$$

Let α_v be the coefficient of cubical expansion. Then, so far as (8.1) applies,

$$(\chi_A)_e \propto 1/V_0 \propto n_v^{-2/3} \propto 1 + \tfrac{2}{3}\alpha_v T. \qquad (8.2)$$

If the temperature factor in (6.9) may be taken as unity

$$\Delta\,(\chi_A)_e/(\chi_A)_e\,\Delta T \doteqdot \tfrac{2}{3}\alpha_v. \qquad (8.3)$$

Table I gives the values of $\Delta\chi_A/\Delta T$ deduced from Sucksmith's results for the metals in the liquid state, and of $(\chi_A)_e$ obtained by allowing for the ionic diamagnetism. The values of α_v are taken from the International Critical Tables.

---

* 'Physica,' vol. 1, p. 979 (1934).

† Stoner, "Magnetism and Matter," ch. xiv (London, 1934).

‡ 'Phil. Mag.,' vol. 2, p. 21 (1926).

TABLE I—TEMPERATURE VARIATION OF SUSCEPTIBILITY OF ALKALI METALS, AND THERMAL EXPANSION

	$\Delta\chi_A/\Delta T \times 10^6$	$(\chi_A)_e \times 10^6$	$\Delta\,(\chi_A)_e/(\chi_A)_e\,\Delta T$	$\frac{2}{3}\alpha_v$
Na..........	$2{\cdot}3 \times 10^{-3}$	18·6	$1{\cdot}2 \times 10^{-4}$	$1{\cdot}8 \times 10^{-4}$
K	$2{\cdot}9 \times 10^{-3}$	32·9	$0{\cdot}9 \times 10^{-4}$	$1{\cdot}9 \times 10^{-4}$
Rb	$8{\cdot}5 \times 10^{-3}$	28·5	$3{\cdot}0 \times 10^{-4}$	$2{\cdot}3 \times 10^{-4}$
Cs	$17{\cdot}7 \times 10^{-3}$	27·0	$6{\cdot}6 \times 10^{-4}$	$2{\cdot}5 \times 10^{-4}$

The values in the last two columns are of the same order of magnitude. More precise agreement could hardly be expected; on the one hand, the value for $\Delta\chi/\Delta T$ can hardly be relied on to more than 10 or 20%, and on the other there is still considerable uncertainty as to the values for χ_A, as judged from the differences in the results of different observers. (For Cs, for example, different investigators give values for $\chi_A \times 10^6$ ranging from -13 to $+29$.) It seems, therefore, that the increase in susceptibility of the alkali metals can be to a large extent accounted for by the change in ε_0 due to thermal expansion, and that there is no immediate experimental contradiction of the formulae derived for free electrons.

One of the most interesting applications of the formulae (6.9) and (6.10) is to elements such as Mn, Pd, and Pt, which have magnetic characteristics which fall between those of elements with practically constant susceptibility (such as the alkali metals, Cu, Ag, and Au), and those for which the variation approaches that given by a Curie law, such as some of the rare earth metals. A general discussion of the magnetic behaviour of metals, however, is outside the purpose of this paper.

SUMMARY

The theoretical expressions previously obtained for the temperature dependence, at low temperatures, of the free electron susceptibility due to electron spin are in disagreement, and there is doubt as to the correctness of the expressions which have been given for the diamagnetic and resultant effects (1). The various formulae which are required in the application of the Fermi-Dirac statistics to the calculation of magnetization are given in a convenient form (2), and calculations are made of the magnetization (M) as a function of H and T for the spin (3) and diamagnetic effects (4) considered separately, and for the resultant effect (5) both at low and high temperatures.

For the resultant effect the following formulae are derived:

For $\varepsilon_0 \gg kT$

$$M = \frac{N\mu^2 H}{\varepsilon_0}\left\{1 - \frac{\pi^2}{12}\left(\frac{kT}{\varepsilon_0}\right)^2 - \tfrac{1}{20}\left(\frac{\mu H}{\varepsilon_0}\right)^2 \ldots\right\}.$$

For $\varepsilon_0 \ll kT$

$$M = \tfrac{2}{3}\,\frac{N\mu^2 H}{kT}\left\{1 - \tfrac{1}{8}\left(\frac{2}{\pi}\right)^{1/2}\left(\frac{\varepsilon_0}{kT}\right)^{3/2} - \tfrac{7}{15}\left(\frac{\mu H}{kT}\right)^2 \ldots\right\}.$$

Here ε_0 is the maximum electron kinetic energy in the completely degenerate state. Under certain conditions these formulae will hold approximately for electrons in metals, with ε_0 as the width of the occupied portion of an unfilled band.

The formulae derived are given in a convenient form for numerical calculations (6). Niessen's treatment is criticized, and the error in his result for the temperature dependence is shown to arise from the incorrect use of a thermodynamic formula (7). The experimental results for metals are briefly considered, and it is shown that the increase in the susceptibility of the alkali metals may be due to the change in the value of ε_0 associated with thermal expansion (8).

The Interaction of Light Nuclei
II—The Binding Energies of the Nuclei H_1^3 and He_2^3

By H. S. W. Massey, Ph.D., Independent Lecturer in Mathematical Physics, Queen's University, Belfast; and C. B. O. Mohr, Ph.D., Exhibition of 1851 Senior Student, Trinity College, Cambridge

(*Communicated by R. H. Fowler, F.R.S.—Received June* 21, 1935)

1—Introduction

The discovery of the light nuclei n_0^1 H_1^2 H_1^3 He_2^3 has provided additional and much-needed material on which to base and test any theory of the structure and interaction on nuclear particles. The properties of these nuclei which are best known are their masses, the latest values of which are the following:

$$n_0^1\ 1\cdot0083 \qquad H_1^2\ 2\cdot0142 \qquad H_1^3\ 3\cdot0161 \qquad He_2^3\ 3\cdot0172,$$

assuming the validity of the mass scheme proposed independently by Oliphant, Kempton, and Rutherford,† and by Bethe‡ (in which the results of the disintegration experiments are found to be consistent if the $He^4 : O^{16}$ ratio is taken as $4\cdot0034 : 16$). From these masses the binding energies of the nuclei, considered as combinations of neutrons and protons, may be obtained. We find, then, that

$$\begin{aligned} H_1^2 &= H_1^1 + n_0^1 - 2\cdot1 \times 10^6\ e\text{-volts},\\ H_1^3 &= H_1^1 + 2n_0^1 - 8\cdot1 \times 10^6\ e\text{-volts},\\ He_2^3 &= 2H_1^1 + n_0^1 - 6\cdot9 \times 10^6\ e\text{-volts}. \end{aligned}$$

The most conspicuous feature of these figures is that the binding energy of both H_1^3 and He_2^3 is considerably greater than twice that of H_1^2, and the question arises as to whether this can be explained without introducing an attractive force between the neutrons in H_1^3 and between the protons in He_2^3. In this paper we attempt to answer the question by applying the variation method to calculate the binding energies of H_1^3 and He_2^3, use being made of all available information bearing on the form of the neutron-proton interaction. It is found that definite results cannot be obtained in this direction, but the calculations would seem to indicate that these additional attractive forces must be introduced, and, in any case, upper limits may be found for their magnitude. Before discussing these calculations it is important to examine the validity of the variation method as applied to H_1^2 by comparing results obtained by its use with exact solutions.

---

† 'Proc. Roy. Soc.,' A, vol. 150, p. 241 (1935).

‡ 'Phys. Rev.,' vol. 47, p. 633 (1935).

2—The Binding Energy of H_1^2 and the Interaction of Neutrons and Protons

The evidence from the scattering of protons by neutrons indicates that the range of interaction between neutron and proton is not greater than 3×10^{-13} cm at most.[†] We therefore assume as reasonable forms for the interaction between a neutron and proton

$$V(r) = -C e^{-2\mu r}; \tag{1}$$

or

$$\left.\begin{aligned} V(r) &= -D, \quad r < a; \\ &= 0, \qquad r > a. \end{aligned}\right\} \tag{2}$$

where $1/(2\mu)$ and a are less than 3×10^{-13} cm. For both cases we can obtain exact spherically symmetrical solutions of the corresponding Schrödinger equation

$$\nabla^2\psi + \frac{4\pi^2 M}{h^2}(E - V)\psi = 0,$$

where M is the mass of either neutron or proton. Thus for the negative energy solutions, corresponding to H_1^2, we have

Case I—

$$\psi = Nr^{-1} J_n\left\{\left(\frac{4\pi^2 M}{\mu^2 h^2} C\right)^{\frac{1}{2}} e^{-\mu r}\right\}, \quad n = \left(\frac{4\pi^2 ME}{\mu^2 h^2}\right)^{\frac{1}{2}},$$

and E is determined by

$$J_n\left\{\left(\frac{4\pi^2 M}{\mu^2 h^2} C\right)^{\frac{1}{2}}\right\} = 0. \tag{3}$$

Case II—

$$\psi = Nr^{-1} \sin \beta r, \qquad r < a, \quad \beta = \left(\frac{4\pi^2 M (D + E)}{h^2}\right)^{\frac{1}{2}},$$

$$= Nr^{-1} \alpha (\alpha^2 + \beta^2)^{-\frac{1}{2}} e^{-\alpha(r-a)}, \quad r > a, \quad \alpha = \left(-\frac{4\pi^2 ME}{h^2}\right)^{\frac{1}{2}},$$

[†] [*Note added in proof, August* 23, 1935.—As no experiments have yet been carried out with *homogeneous* neutron beams this evidence is perhaps not so definite as would appear at first sight.]

and E is determined by

$$\tan \beta a = -\beta/\alpha. \tag{4}$$

If we put E equal to $-2 \cdot 1 \times 10^6$ *e*-volts, the observed binding energy of H_1^2 the equations (3) and (4) relate the constants C, μ of (1) and D, *a* of (2) respectively. In fig. 1 the relations between the constants are illustrated. The only additional restriction on this relation which we can impose is that $1/\mu$ and *a* must be less than 3×10^{-13} cm.

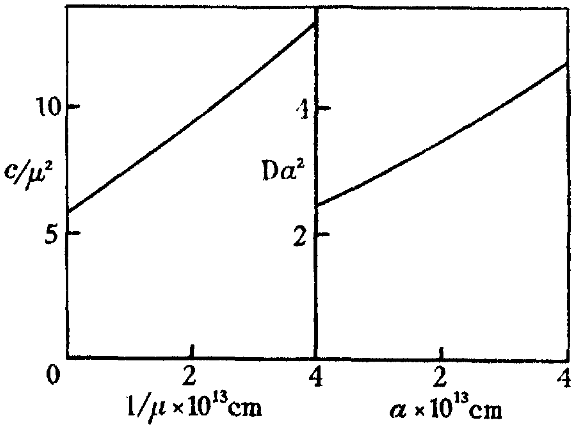


FIG. 1.

We now compare the results which would be obtained with the variation method, using various trial wave functions. The variation method gives

$$-\frac{4\pi^2 ME}{h^2} = \frac{\int \psi^* \left(\nabla^2 \psi - \frac{4\pi^2 M}{h^2} V \psi \right) dv}{\int \psi^* \psi \, dv} = I, \tag{5}$$

where $\delta I = 0$. In applying this method, reasonable trial forms for ψ are taken containing one or more parameters which are varied to give $\delta I = 0$ with respect to their variations, and hence a value for E. To test the accuracy of the results obtained in this way, the form (1) was taken for V (*r*) and calculations carried out for different values of μ and the corresponding values of C which give the observed binding energy of H_1^2, according to the exact solution (3).

The most natural form of wave function to try first is the simple one†

$$\psi = e^{-\alpha r}. \tag{6}$$

---

† The value of α here bears no relation to the value of α in (4).

The binding energies obtained with this function are compared with the true values for various μ in fig. 2, and it will be seen that for $\mu > 1$ this function gives extremely unsatisfactory results. The reason for this is clear. As μ increases the true wave function has increasingly marked concentration in the region of small r but falls off for large r always at the same rate, viz.,

$$\psi \to r^{-1} \exp\left\{-\left(\frac{4\pi^2 ME}{h^2}\right)^{\frac{1}{2}} r\right\}, \quad \text{as} \quad r \to \infty .$$

The single exponential form (6) represents a sort of " mean " of these rapidly- and slowly-varying forms and becomes less and less satisfactory as the variation of the wave function between small and large r increases.

The function

$$\psi = e^{-\alpha r^2} \tag{7}$$

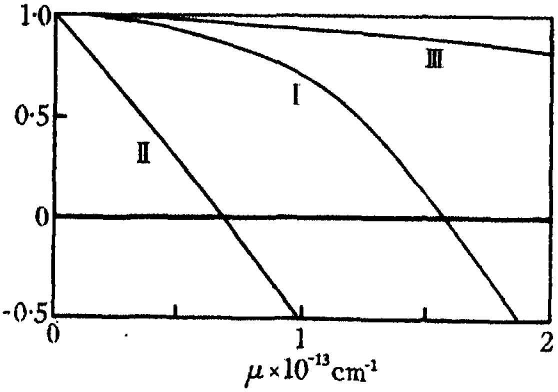


FIG. 2—Ratio of binding energy of H_1^2 calculated by the variation method to the true value, using

I $\psi = e^{-\alpha r}$, II $\psi = e^{-\alpha r^2}$, III $\psi = r^{-1}(e^{-\alpha r} - e^{-\beta r})$.

was also tried and found to give much less satisfactory results for large μ, see fig. 2. It gives the steep maximum near $r = 0$, but not the correct asymptotic form. Eventually the function

$$\psi = r^{-1}(e^{-\alpha r} - e^{-\beta r}) \tag{8}$$

containing two parameters, α and β, was used. The correct asymptotic form may be obtained with the parameter α, and β can be adjusted to give the correct form near the origin. Results obtained with this wave function are illustrated in fig. 2 and it is clear that, although by no means very accurate for large μ, the exact value of the binding energy is approximated to much more closely than with the forms (6) and (7). It was not found possible to obtain a simple wave function giving more accurate results than this one, and it was accordingly used whenever possible in the further calculations on H_1^3 and He_2^3 where variation methods must be relied on.

It is of interest to notice here how different the situation is from the corresponding problem of calculating the binding energies of atoms. This difference is due to the small range of the interaction energy in nuclear problems, in sharp contrast to the slowly-varying Coulomb field. The solutions of nuclear problems must follow on similar lines to those associated with negative atomic ions, and it is already known that the latter present difficulties which do not arise for neutral atoms. For this reason we have given a rather detailed account of the application of the variation method to H_1^2, to reveal the peculiarities of behaviour inherent in short range interactions.

3—THE BINDING ENERGY OF H_1^3

If we choose as co-ordinates the vectors $\mathbf{r}_1$, $\mathbf{r}_2$ joining the proton to the two neutrons and position vector $\mathbf{R}$ of the centre of mass we have the wave equation for H_1^3, after separation of the irrelevant terms involving $\mathbf{R}$

$$
\begin{aligned}
H\Psi \equiv \nabla_1^2\Psi + \nabla_2^2\Psi + \nabla_1 \cdot \nabla_2\Psi \\
- \frac{4\pi^2 M}{h^2}\{V(r_1) + V(r_2) + W(|\mathbf{r}_1 - \mathbf{r}_2|)\}\Psi = - \frac{4\pi^2 M}{h^2} E, \qquad (9)
\end{aligned}
$$

where M is the mass of a neutron or proton, V (r) the interaction between a neutron and proton, and W (r) that between two neutrons. This wave equation is derived on the assumption of no three-body forces and absence of effects due to internal co-ordinates of neutron or proton. Let us now suppose $W = O$ and proceed to evaluate E. If we neglect the term $\nabla_1 . \nabla_2 \Psi$ involving mixed differentials the binding energy is obviously twice that of H_1^2, and, moreover, since the variation method gives

$$
- \frac{4\pi^2 M}{h^2} E = I = \frac{\iint \Psi^* H \Psi \, dv_1 \, dv_2}{\iint \Psi^* \Psi \, dv_1 \, dv_2}, \qquad (10)
$$

with the parameters of Ψ chosen so $\delta I = 0$ with respect to their variations, we see that any trial function which involves only the distances r_1, r_2 introduces no contribution from $\nabla_1 . \nabla_2$, for

$$
\begin{aligned}
&\iint \Psi^* (r_1, r_2) \nabla_1 . \nabla_2 \Psi (r_1, r_2) \, dv_1 \, dv_2 \\
&\quad = \iint \{\cos\theta_1 \cos\theta_2 + \sin\theta_1 \sin\theta_2 \cos(\phi_1 - \phi_2)\} f(r_1, r_2) \, dv_1 \, dv_2 \\
&\quad = 0.
\end{aligned}
$$

Now we have already pointed out in the introduction that the observed binding energy of H_1^3 is considerably greater than twice that of H_1^2 so, if this is to be explained with $W = 0$, we must look to terms arising from $\nabla_1 . \nabla_2 \Psi$ to give the extra energy. It is very difficult to make any *a priori* estimate of the amount that this term is likely to contribute. We might attempt to estimate its possibilities by a second order perturbation calculation. The unperturbed wave function to be taken is

$$\Psi (r_1, r_2) = \psi_0 (r_1) \psi_0 (r_2),$$

where $\psi_0 (r)$ is the wave function corresponding to the bound state of H_1^2. The second order perturbation energy is given approximately by[†]

$$E_2 = \frac{h^4 E_0}{16\pi^4 M^2} \iint \Psi (r_1, r_2) (\nabla_1 . \nabla_2)^2 \Psi (r_1, r_2) \, dv_1 \, dv_2,$$

where E_0 is the zero order energy. Taking the form (8) for $\psi (r)$ we find

$$E_2 = \frac{h^2}{4\pi^2 M} \frac{\alpha\beta}{6},$$

and so tends to infinity as β tends to infinity (for α remains always nearly equal to $0 \cdot 22 \times 10^{13}$ cm^{-1}), *i.e.*, as the range of interaction tends to zero. Even for values of $1/\mu$ as small as 10^{-13} cm, E_2 is comparable with E_0. This suggests that it might be possible to explain the binding energy of H_1^3 without invoking the interaction W between the neutrons. The extreme sensitiveness of the binding energy for these short range fields to small changes in the field constants also makes this conceivable. Unfortunately, there appears at the moment to be only one way of proceeding further, and that is by trial, using the variation method. One can never be certain that this method does not give a binding energy much smaller than the true value, although it can never give a greater value.

In applying the variation method to make use of the term in $\nabla_1 . \nabla_2 \Psi$ we must use a wave function

$$\Psi = \psi (r_1) \psi (r_2) \{1 + f (r_1, r_2, \Theta)\}, \tag{11}$$

where Θ is the angle between $\mathbf{r}_1$ and $\mathbf{r}_2$. To obtain a reasonable value for the total binding energy it is necessary that $\psi (r_1)$, $\psi (r_2)$ should approximate closely to accurate wave functions for the H_1^2 problem as the binding energy obtained is markedly diminished if this is not so. The term $f (r_1, r_2, \Theta)$ represents the effect of $\nabla_1 . \nabla_2 \Psi$. Now for maximum potential energy in the system $\Theta = 0$, but then the kinetic energy is also a maximum and the total energy is actually a minimum for $\Theta > 0$. It is natural to try initially forms like

---

[†] Lennard-Jones, 'Proc. Roy. Soc.,' A, vol. 129, p. 598 (1930).

$$f(r_1, r_2, \Theta) = \gamma \cos \Theta \begin{cases} r_1 r_2, \\ 1, \\ r_1 r_2^2 + r_2 r_1^2, \end{cases} \tag{12}$$

with $\psi(r_1)$, $\psi(r_2)$ of the same form as for H_1^2 but with adjustable constants. For $\mu < 0 \cdot 5 \times 10^{13}$ cm^{-1} it is possible to use the simple form (6), but for $\mu > 0 \cdot 5 \times 10^{13}$ cm^{-1} the more accurate form (8) must be used to obtain a total binding energy of appreciable magnitude. A large number of different forms for Ψ' were tried, and the best was

$$\Psi' = r_1^{-1} r_2^{-1} (e^{-\alpha r_1} - e^{-\beta r_1})(e^{-\alpha r_2} - e^{-\beta r_2})(1 + \gamma r_1 r_2 \cos \Theta), \tag{13}$$

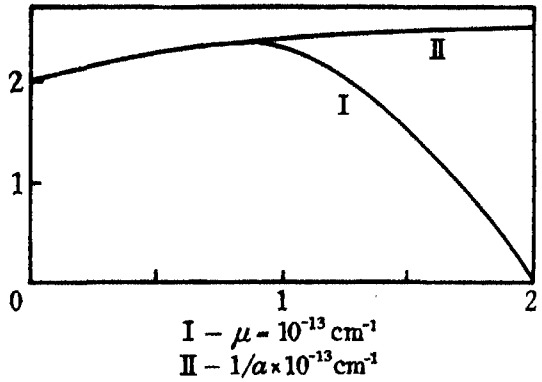


FIG. 3—Ratio of calculated binding energy of H_1^3 to that of H_1^2, using I—the form for Ψ' given in relation (13), II—"square hole" type of *n-p* interaction of radius *a*, and assuming no force between the neutrons.

for which the calculated binding energies for H_1^3 for different values of μ are illustrated in fig. 3. It will be seen that the maximum value is only 2·4 times the true binding energy of H_1^2, although it is as much as 2·9 times the binding energy of H_1^2 given by the trial function (6) (even this is less than the observed ratio 3·9). From this last result it is clear that a considerable improvement could be obtained for large values of μ if a more accurate wave function for H_1^2 could be used. The results obtained using (13) showed that the optimum values of the constants α and β are little affected by the introduction of the terms calculated for $\nabla_1 . \nabla_2 \Psi'$, and so calculations were next carried out using the interaction energy (2) and taking $\psi(r_1)$, $\psi(r_2)$ as exact solutions of the equation for the bound state of H_1^2 for this interaction energy. This led to a slight improvement when $f(r_1, r_2, \Theta)$ was taken as $1 + \gamma r_1 r_2 \cos \Theta$, and the results are illustrated[†] in fig. 3. It is seen that the curve tends asymptotically to a value of 2·5 times the binding energy of H_1^2. To improve the result still further, $f(r_1, r_2, \Theta)$ was expanded in the form

$$\sum_n \gamma_n r_1{}^n r_2{}^n P_n(\cos \Theta),$$

but the contribution from terms with $n > 1$ was found inappreciable. Introduction of different powers of r_1, r_2 also gave no improvement, and the same lack of success occurred with the form $\gamma r_1 r_2 \sin \Theta$. It is difficult to see what other form[‡] could be used which would be at all likely to add the additional 3×10^6 *e*-volts to the binding energy.[§] On these grounds it is reasonable but by no means definite to assume that the actual value of the binding energy when $W = 0$ is not appreciably greater than the best value obtained in the way described above, *i.e.*, a binding energy 2·5 times that of H_1^2 for a neutron-proton interaction of form (2) with $a < 0 \cdot 7 \times 10^{-13}$ cm. In this connexion it is important to notice that we have taken account of the effect of the short range interaction on the form of the spherically symmetrical solutions for H_1^2, and have adjusted the constants in the interaction energy function to give the observed binding energy of H_1^2 by *exact* methods and not by the variation method (which we have shown in § 2 does not give the correct constants except under certain conditions).

Proceeding on the assumption that a finite value of W is required, let us see what restrictions we can place on its magnitude.

4—RESTRICTIONS ON THE INTERACTION ENERGY BETWEEN NEUTRONS

When we take into account the finite value of the interaction energy W it is convenient to transform the wave equation into a new set of coordinates, the distances r_1, r_2 between the proton and each neutron, and the distance u between the neutrons and the angles defining the orientation of the triangle marked out by the particles. As the wave function for the ground state will not involve the angles of orientation, we are left with the wave equation

---

[†] It will be seen, by reference to the various figures, that the forms (1) and (2) always give comparable results, if one takes $1/\mu$ to correspond approximately to a.

[‡] Calculations were also carried out using the wave function $\psi = e^{-\alpha\rho-\beta u}$, where ρ is the distance of the proton from the centre of mass of the neutrons, u the distance between the neutrons, without yielding any improvement.

[§] [*Note added in proof, August* 23, 1923. While this paper has been in press.— L. H. Thomas has published ('Phys. Rev.,' vol. 47, p. 903 (1935)), an investigation of H_1^3, using a variation method. Taking a very complicated trial function, he finds that the calculated energy of H_1^3 can be indefinitely increased if the range of interaction of neutron and proton is indefinitely diminished. For larger ranges of interaction, however, the results he obtains are substantially in agreement with those reported above. The conclusions of §§ 4 and 5 remain valid under these conditions.]

$$\frac{1}{r_1^2}\frac{\partial}{\partial r_1}\left(r_1^2\frac{\partial\Psi}{\partial r_1}\right)+\frac{1}{r_2^2}\frac{\partial}{\partial r_2}\left(r_2^2\frac{\partial\Psi}{\partial r^2}\right)+\frac{1}{u^2}\frac{\partial}{\partial u}\left(u^2\frac{\partial\Psi}{\partial u}\right)$$

$$+\frac{r_2^2+u^2-r_1^2}{2r_2u}\frac{\partial^2\Psi}{\partial r_2\partial u}+\frac{r_1^2+u^2-r_2^2}{2r_1u}\frac{\partial^2\Psi}{\partial r_2\partial u}+\frac{r_1^2+r_2^2-u^2}{2r_1r_2}\frac{\partial^2\Psi}{\partial r_1\partial r_2}$$

$$-\frac{4\pi^2\mathrm{M}}{h^2}(\mathrm{V}(r_1)+\mathrm{V}(r_2)+\mathrm{W}(u))=-\frac{4\pi^2\mathrm{M}}{h^2}\mathrm{E}.$$

We now have to consider the possible values of W. It is natural to expect that W (u) will have a range comparable with that of V (r) and also will be of the same order of magnitude. Useful information can then be obtained by taking for W, corresponding to cases (1) and (2) of § 2 for V, either

$$\mathrm{W}(u)=-\mathrm{C}'e^{-2\mu' u};\tag{14}$$

or

$$\left.\begin{aligned}\mathrm{W}(u)&=-\mathrm{D}', & u&<a';\\ &=0, & u&>a'.\end{aligned}\right\}.\tag{15}$$

To avoid too large a number of variables we set $\mu'=\mu$ (or $a'=a$) and calculate the additional binding energy as a function of C′ (or D′) for different values of μ (or a).

Using the form (14) we are not able to carry the calculations out for large μ, for in this case it is necessary to use trial functions of the form (8), and with the form (14) for W the required integrations cannot be carried out. However, for $\mu < 0\cdot7\times10^{13}$ cm, the trial function

$$\Psi(r_1,r_2,u)=e^{-a(r_1+r_2)\beta u}\tag{16}$$

may be used to give sufficiently accurate results. The addition to the binding energy due to W for a fixed μ (and hence μ') is practically proportional to C and increases (for C′ = C) as μ increases, though the trial function soon becomes so bad as to give a negative total binding energy.

For values of $1/\mu$ for which we can trust the trial function (16) (between $1\cdot5$ and 3×10^{-13} cm), and which also fit in with collision experiments (say between $1\cdot0$ and 3×10^{-13} cm), it is found that the total binding energy due to V (r_1), V (r_2), and W (u) together would markedly exceed that of H_1^3 if C′ were as great as C, and the indications are that this holds also for all larger values of μ. To confirm this it is necessary to use the form (3) for V and (15) for W, and take as trial function

$$\Psi = \psi(r_1)\,\psi(r_2),$$

where $\psi(r)$ is the exact solution for the bound state of H_1^2. Using this we find values of the extra binding energy for $D^1 = D$ and $a^1 = a$. These values are illustrated in fig. 4 for different values of $1/\mu$ and a. Since the extra energy required is 3×10^6 *e*-volts, we can say definitely that C′ (or D′) cannot be much greater than $\frac{1}{3}$ C (or $\frac{1}{3}$ D), so that the magnitude of the interaction between two neutrons must be less than one-third of that between a neutron and proton if both have the same range. This excludes the possibility of existence of a stable nucleus n_2^0, for this requires an interaction energy which, for a given range, is not appreciably less than that between a neutron and a proton.

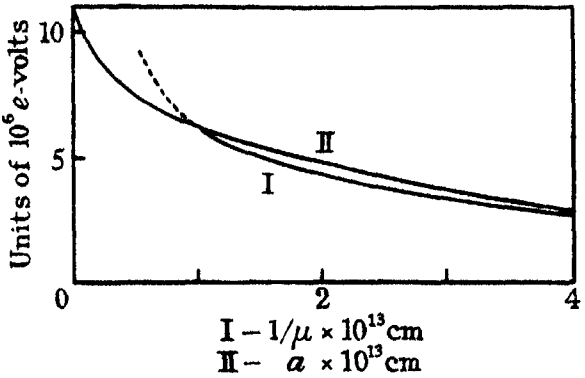


FIG. 4—Contribution to the binding energy of H_1^3 which would be due to an attraction between neutrons equal to that between neutron and proton, for different " sizes " of neutron.

The sensitiveness of the binding energy to interaction between the neutrons is thus very marked. Although the binding energy of H_1^3 is roughly four times that of H_1^2, this can be secured by taking the neutron-neutron interaction only $\frac{1}{3}$ of the neutron-proton. It is this sensitiveness which makes it still possible that a function might be found† which will give the observed binding energy of H_1^3 by the variation method without recourse to any neutron-neutron interaction, but we can at least be definite in saying that this latter interaction, if it exists, must be of a considerably smaller magnitude than that between a neutron and proton.

---

† *Cf.* Footnote §, p. 700.

5—THE BINDING ENERGY OF He_2^3 AND THE INTERACTION BETWEEN PROTONS

We now turn to the nucleus of He_2^3, composed of two protons and a neutron. The binding energy here is still considerably greater than twice that of H_1^2 and, in fact, is greater than that calculated for H_1^3 in § 3, neglecting neutron-neutron interaction. As the Coulomb repulsion between the protons, supposing it to hold down to small distances, would reduce the binding energy by from 0·5 to 1·0 million volts we see that, if neutron-neutron interaction is necessary to explain the mass of H_1^3, a failure of the Coulomb law of force between protons at small distances is required to explain the mass of He_2^3.

We may apply the variation method to set an upper limit to the extent of this deviation and, assuming reasonable accuracy of the method, we can estimate the actual magnitude of the deviations. If U (u) is the interaction energy between two protons at distance u apart we take, corresponding to (14) and (15),

$$\mathrm{U}\,(u) = e^2 \,.\, \frac{1 - e^{-2\mu' u}}{u} - \mathrm{C}'\, e^{-2\mu' u}; \tag{17}$$

or

$$\left.\begin{aligned} \mathrm{U}\,(u) &= -\,\mathrm{D}', & u &< a'; \\ &= \frac{e^2}{u}, & u &> a'. \end{aligned}\right\}, \tag{18}$$

and proceed as in § 4 for H_1^3, putting $\mu' = \mu$ (or $a' = a$). For large μ (17) must be used as in § 4, and for small results obtained by using (17) and (18) agree. These results are illustrated in fig. 5, where the extra binding energy due to U obtained in this way is plotted as a function of $1/\mu$ (or a) for C′ equal to C (or D′ = D). As the required additional energy is not greater than 3×10^6 e-volts, we see that C′ (or D′) cannot be greater than 1/3 C (or 1/3 D) for any value of $1/\mu$ (or a). For such values of C′ (or D′) no stable combination of two protons could exist.

Fortunately, there is the additional possibility of investigating any failure of the Coulomb law of force between protons by collision experiments, and White† has recently reported results which indicate anomalous scattering of 750 k volt protons in hydrogen. His values are as yet not complete enough to lead to any definite conclusions, but they can be explained by supposing that the anomalous field produces a phase shift of roughly 45° in the zero order incident wave (this phase shift being measured relative to the wave which would exist in a purely Coulomb field). This is of the same order of magnitude as the phase shift which would be produced by an anomalous field of the type obtained above with C′ or D′ of the same order of magnitude as our calculated values. Further experimental evidence in this connexion would clearly be most valuable.

---

† 'Phys. Rev.,' vol. 47, p. 573 (1935).

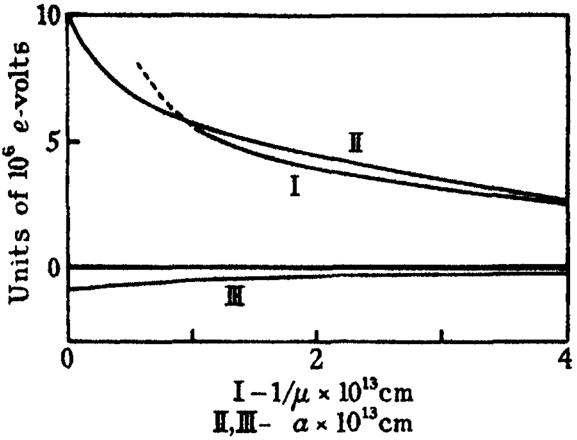


FIG. 5—Contribution to the binding energy of He_2^3 which would be due to an attraction at close distances between protons equal to that between neutron and proton, for different " sizes " of proton. Curve III represents the effect of the Coulomb repulsion alone, taken to exist at distances greater than a.

6—CONCLUDING REMARKS

We may summarize the conclusions arrived at as follows:

(*a*) The variation method must be applied with caution in determining binding energies due to short range interactions.

(*b*) It is unlikely, but not impossible,† that the binding energy of H_1^3 and He_2^3 can be explained without introducing attractive forces between the neutrons on the one hand and between the protons on the other.

(*c*) The attraction between two neutrons must be much less than that between a neutron and a proton, so the existence of a stable n_0^2 nucleus is precluded.

(*d*) The anomalous attraction between two protons at small distances must also be much less than that between a neutron and a proton, and cannot be sufficient to produce a stable nucleus He_2^2.

It is important to remember that we have taken no account of the possibility of the occurrence of three-body forces in the nuclei concerned. The existence of such forces might invalidate conclusions (*c*) and (*d*).

---

† *Cf.* Footnote §, p. 700.

7—SUMMARY

The accuracy of the variation method in dealing with short range interactions has been investigated by comparison with exact solutions for H_1^2, and a great variety of trial functions have been employed in calculating the binding energy of H_1^3 on the assumption of zero interaction between the neutrons. No function was found which gave a binding energy greater than $5 \cdot 4 \times 10^6$ *e*-volts (observed value is $8 \cdot 1 \times 10^6$ *e*-volts). Introducing an attraction between the neutrons, it has been shown that the observed binding energy can be obtained if this attraction is even smaller than one-third of that between neutron and proton, if the ranges of the interaction in the two cases are comparable. This precludes the possibility of the existence of a stable nucleus n_0^2. Similar calculations for He_2^3 show that an anomalous attraction between two protons at small distances of the same order as that between two neutrons will give the observed binding energy, and consequently the nucleus He_2^2 must be unstable.

The Diffusion of Gases through Metals
II—Diffusion of Hydrogen through Aluminium

By C. J. SMITHELLS, D.Sc., and C. E. RANSLEY, B.Sc. (Research Staff of the General Electric Co., Ltd., Wembley)

(*Communicated by R. H. Fowler, F.R.S.—Received June* 21, 1935)

The only direct attempt to determine the rate of diffusion of hydrogen through solid aluminium, of which we are aware, is described in a paper by Deming and Hendricks,[*] in which they conclude that aluminium is impermeable to hydrogen at temperatures up to 550° C. Very slow rates of diffusion would not have been detected by the method employed.

The solubility of hydrogen in aluminium has been measured by Iwasé,[†] Röntgen and Braun,[‡] and Röntgen and Möller.[§] No solubility was detected in solid aluminium, but molten aluminium dissolves about 0·1 to 0·2 cc of hydrogen per 100 gm of metal at the melting-point, and the solubility increases with temperature.

It is well known that molten aluminium in contact with the ordinary atmosphere decomposes water vapour with the formation of aluminium oxide, and that much of the hydrogen liberated is absorbed by the metal. On freezing, the greater part of this hydrogen is liberated in the form of bubbles, some of which remain entrapped in the solid metal. As a result of this, commercial aluminium usually contains an appreciable amount of gas, 80% of which is hydrogen. An average analysis shows about 14 cc of gas per 100 gm of metal.[||] Hanson and Slater[¶] concluded from indirect evidence that during atmospheric corrosion of aluminium the nascent hydrogen produced is absorbed by the metal, and if this is true the gas must be soluble in the solid metal.

In an earlier paper[**] we have shown that the diffusion of a gas into a solid metal depends on adsorption on the surface and may be markedly influenced by the presence of an oxide film. Pilling and Bedworth[††] have shown that a film of aluminium oxide 0·00002 cm thick on aluminium is impervious to oxygen. Since aluminium in contact with the atmosphere is always covered by a film of oxide it seemed possible that the failure to establish either diffusion or solid solubility of hydrogen might be due to the action of this film in preventing diffusion of the gas into, or out of, the metal.

---

[*] 'J. Amer. Chem. Soc.,' vol. 45, p. 2857 (1923).

[†] 'Sci. Rep. Tôhoku Imp. Univ.,' vol. 15, p. 531 (1926).

[‡] 'Metallwirtschaft,' vol. 11, p. 459 (1932).

[§] 'Metallwirtschaft,' vol. 11, pp. 685, 697 (1932).

[||] Röntgen and Braun (*loc. cit.*).

[¶] 'J. Inst. Metals,' vol. 46, p. 187 (1931).

[**] 'Proc. Roy. Soc.,' A, vol. 150, p. 172 (1935).

[††] 'J. Inst. Metals,' vol. 29, p. 573 (1923).

We have therefore examined the diffusion of hydrogen through aluminium in the normally tarnished condition, after anodic oxidation, and with the oxide film removed by abrasion *in vacuo*. Preliminary results already published* led us to the conclusion that hydrogen diffuses through aluminium. These results are confirmed, and it is shown that the rate of diffusion is dependent on the state of the surface.

APPARATUS AND METHOD

The apparatus used was substantially that described in the earlier paper except for modifications introduced so that both the inner and outer surfaces of the aluminium tube could be vigorously scraped *in vacuo* without removing the tube from the diffusion apparatus. Only the modifications to the apparatus need be described.

Material—Commercial aluminium bar 1 inch in diameter was drilled and drawn into tubing having an external diameter of 3·2 mm and a wall thickness of 0·45 mm. The metal had the following analysis:

	%
Copper	0·03
Iron	0·90
Manganese	0·03
Silicon	0·4
Zinc	0·17

A piece of tube 30 cm long was closed at one end by welding, and the other end was soldered to a copper-to-glass joint leading to the analytical system.

Surface Treatment—When aluminium is exposed to the air a transparent but protective film of oxide forms rapidly on the surface. The tube when introduced into the diffusion apparatus had been exposed for some days and was therefore in this condition. The only practical method of removing this film is by mechanical abrasion, and arrangements were made for scraping both the inside and outside surfaces of the tube *in vacuo* in the diffusion apparatus. Scratch brushes of fine steel wire bearing with considerable pressure on both the inner and outer surfaces were attached by nickel tubes to blocks of iron which could be moved horizontally by means of a powerful solenoid, so that either surface could be scratched independently. This method of controlling mechanical movement in a vacuum was described in the previous paper with respect to the exploring thermocouple. The brushes were found to be very effective, producing pronounced scratches over the surface of the tube, and a considerable quantity of abraded matter collected in the bottom of the diffusion bulb.

---

* 'Nature,' vol. 135, p. 548 (1935).

Temperature Control—The presence of the scraper made it difficult to attach current leads as before, and the tube was therefore heated indirectly by means of a furnace, which consisted of a tungsten helix between two concentric quartz tubes. The temperature distribution along the tube, measured as before with an exploring thermocouple, is less satisfactory with this method of heating. The equivalent length of tube at constant temperature was estimated graphically, using the temperature coefficient of diffusion determined as described later in this paper. The effective hot length of tube varied from 3 to 4 cm in different experiments.

Preliminary Out-gassing—Before carrying out diffusion measurements the tube was thoroughly out-gassed at 605° C, and the glass work baked out with both sides of the apparatus connected to the vacuum pumps. The out-gassing was usually complete after 12 hours, and was repeated each time the tube was exposed to the atmosphere.

EXPERIMENTAL RESULTS

Effect of Surface Treatment—A series of diffusion measurements were made on the same tube subjected to various surface treatments. All these experiments were made with the tube at 580° C (853° K) and a hydrogen pressure of 29 cm, under which conditions sufficient hydrogen diffused during half an hour to enable an accurate estimate of the rate to be made in that time. The details of these separate runs are given below and the results are summarized in fig. 1. As before, the rate of diffusion D is expressed as the volume of gas in cubic centimetres at N.T.P. diffusing per second through 1 sq. cm of surface of 1 mm thickness.

Run 1—The tube was run in hydrogen for 3 hours before any observations were made. The rate was then $2 \cdot 0 \times 10^{-7}$ but appeared to be diminishing slightly with time.

Run 2—The tube was allowed to cool and, without altering any other condition, the outer surface was vigorously scratched. The tube was again heated to 580° C, and at the end of the first hour the rate ($3{\cdot}8 \times 10^{-7}$) was appreciably higher than before, but was definitely decaying.

Run 3—The outer surface was scratched again and the average rate of diffusion determined during successive ½-hour periods. The initial rate was now $4{\cdot}2 \times 10^{-7}$, but fell during 5 hours to a fairly constant value at $1{\cdot}4 \times 10^{-7}$.

Run 4—The outer surface was scratched again, producing an initial rate of $4{\cdot}0 \times 10^{-7}$, which fell in 7 hours to $0{\cdot}9 \times 10^{-7}$.

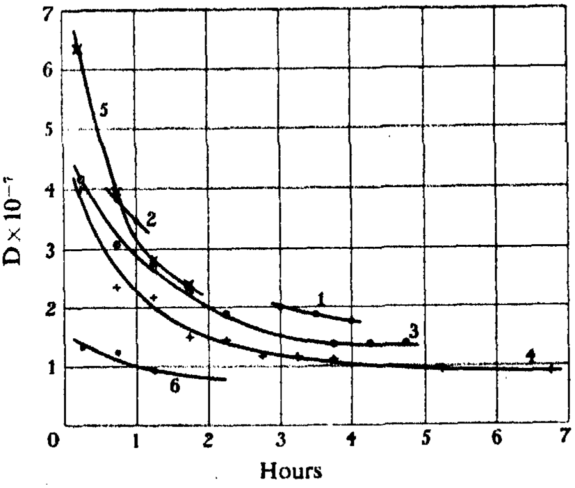


FIG. 1—Effect of surface treatment on rate of diffusion. Run 1, exposed to air; run 2, outer surface scratched; run 3, outer surface scratched again; run 4, outer surface scratched again; run 5, inner and outer surfaces scratched; run 6, outside anodically oxidized.

Run 5—Both the inner and outer surfaces of the tube were scratched. The average rate during the first half-hour was now $6{\cdot}3 \times 10^{-7}$, and from the rapid rate of decay it would seem that the rate during the first few minutes must have been nearly $10{\cdot}0 \times 10^{-7}$.

From these experiments it appeared that hydrogen diffuses quite readily through aluminium at 580° C, the rate being of the same order as we have found for copper at this temperature. It is apparent, however, that the rate may vary by a factor of at least 10 to 1 according to the condition of the surface. Whilst the initial rates in the different runs vary considerably, the curves of fig. 1 indicate that after about 10 hours the surface returns to a definite state, for which the rate is about $0{\cdot}8 \times 10^{-7}$. It seems almost certain that this condition corresponds to a definite degree of oxidation of the surface. Although the hydrogen was pure and dry, being introduced by diffusion through a palladium tube, some of it would be converted by contact with the hot tube into atomic hydrogen, which can reduce the oxidized parts of nickel wires, etc., even in the cold. The water vapour so formed would oxidize the aluminium. It seemed of interest, therefore, to determine the rate of diffusion through aluminium covered with a definitely protective oxide film such as that produced by anodic oxidation.

Run 6—The tube was removed from the apparatus, carefully polished and anodically oxidized in a 3% chromic acid solution. After the usual out-gassing the initial rate of diffusion was found to be $1 \cdot 33 \times 10^{-7}$, and this fell during a period of 2½ hours to about $0 \cdot 9 \times 10^{-7}$. These rates are of the same order as the final values obtained in the previous runs, and support the conclusion that the decay is due to oxidation of the surface.

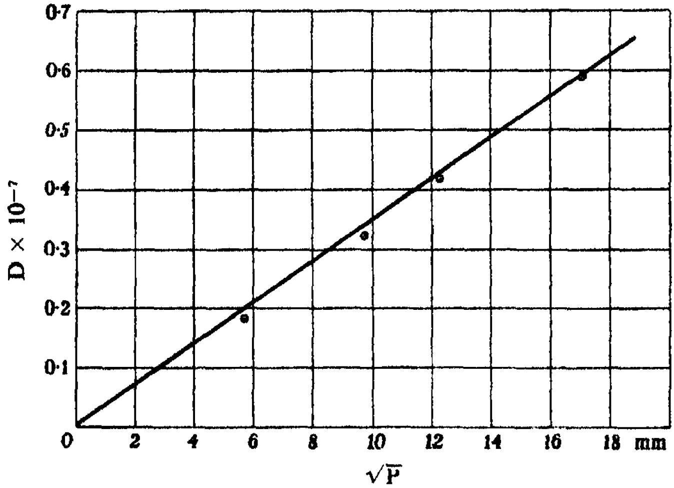


FIG. 2—Effect of pressure on rate of diffusion at 831° K.

Effect of Pressure—The effect of pressure on the rate of diffusion at 558° C was determined for a tube in the fully decayed condition ($D = 0 \cdot 8 \times 10^{-7}$ at 580° C and 29 cm pressure). The results, which are given in Table I and plotted in fig. 2, are represented by the usual equation

$$D = k \sqrt{P}$$

within the limits of experimental error. The rate of diffusion was too slow to enable accurate measurements to be made at very low pressures, and it is not possible to say whether the usual deviation* from the square root law at low pressures takes place.

Effect of Temperature—The effect of temperature on the rate of diffusion was determined for a tube having the high rate of diffusion characteristic of a freshly scraped surface, and also for the anodically oxidized tube.

The high rate was obtained with the same tube used for the previous measurements, carefully polished before assembling, but not otherwise specially treated. On this occasion the tube maintained a high rate ($5 \cdot 0 \times 10^{-7}$ at 580° C and 29 cm pressure) for 7 hours without showing any marked decay, although the experimental conditions were normal. The results are given in Table II, and in fig. 3, log D is plotted against 1/T. It is clear that the usual exponential equation

TABLE I—EFFECT OF PRESSURE ON RATE OF DIFFUSION AT 558° C (831° K)

Pressure mm	Rate D
32·5	$0 \cdot 18 \times 10^{-7}$
95	0·32
150	0·42
290	0·59

$$D = ke^{-b/T}$$

represents the effect of temperature on the rate of diffusion. The value of b given by the slope of the line in fig. 3 is 15,600. This is in close agreement with the value of 15,400 found previously† for a different tube. The temperature coefficient is therefore much higher than for any other metal-hydrogen system for which data are available.

It was difficult to obtain consistent results with the anodically oxidized tube owing to the very low rate of diffusion at any but the highest temperatures.

The observations plotted in fig. 3, and marked in the order they were made, indicate a slightly higher temperature coefficient for the anodically oxidized surface. Since the actual rate of diffusion of hydrogen through aluminium varies so much with the condition of the surface, it is not possible to give an absolute value to the diffusion coefficient k in the equation

$$D = \frac{k\sqrt{P}}{d} e^{-b/T}.$$

---

* Smithells and Ransley, 'Proc. Roy. Soc.,' A, vol. 150, p. 172 (1935).

† Smithells and Ransley, 'Nature,' vol. 135, p. 548 (1935). In this note the value was given as 14,000, owing to an arithmetical error, instead of 15,400.

TABLE II—EFFECT OF TEMPERATURE ON THE RATE OF DIFFUSION AT 29 CM PRESSURE

Temperature °K	Rate D
782	$1 \cdot 09 \times 10^{-7}$
793	1·17
805	1·75
818	2·33
842	4·09
865	5·92
878	9·70

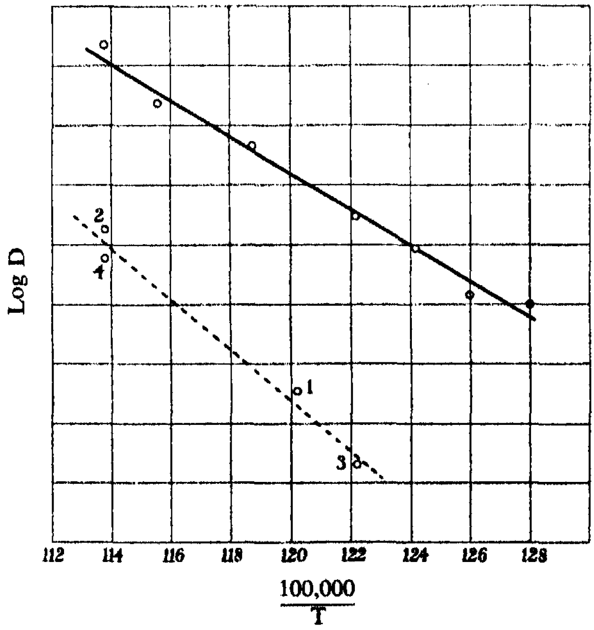


FIG. 3—Effect of temperature on the rate of diffusion at 29 cm. —— after exposure to air; - - - after anodic oxidation.

The value found in our experiments varies from 3·3 for a freshly scraped surface to 0·42 for an anodically oxidized surface.

DISCUSSION

These experiments show that hydrogen diffuses through solid aluminium. The usual diffusion equation represents the effects of temperature and pressure on the rate of diffusion, and we therefore conclude that the hydrogen is dissociated at the metal surface. Even under the best conditions it is difficult, if not impossible, to avoid contamination of the surface, probably by oxygen. This reduces the rate of diffusion both into and out of the metal. The very high temperature coefficient of diffusion, in conjunction with this poisoning effect, results in extremely slow rates of diffusion at temperatures more than 100° below the melting-point of the metal. The fact that hydrogen does diffuse through aluminium shows that the gas must have some solubility in the solid metal. The failure of previous authors to detect this solubility is probably due to the slow rate of diffusion under ordinary experimental conditions.

SUMMARY

It is found that hydrogen diffuses through aluminium at a measurable rate at temperatures above 400° C. The rate of diffusion depends on the state of the surface. The highest rate is obtained with a surface freshly scraped in hydrogen, but it falls off rapidly and after some hours reaches a steady value which is only about one-tenth of the initial rate. This is attributed to contamination of the surface by oxygen, since this low rate is obtained with an anodically oxidized surface.

The effects of temperature and pressure are satisfactorily represented by Richardson's equation

$$D = k\sqrt{P}e^{-b/T}.$$

The value of b for the freshly scraped surface is 15,600 and for the anodically oxidized surface about 21,500. The value of k varies from 3·3 to 0·42 for different states of the surface.

The Energy of Disintegration of Radio-Phosphorus [P^{30}]

By C. D. ELLIS, F.R.S., and W. J. HENDERSON, Ph.D. (Exhibition of 1851 Student, Queen's University, Kingston, Canada)

(*Received October* 23, 1935)

1—INTRODUCTION

The exact masses of the nuclei are quantities of great interest depending directly on the forces of cohesion between the nuclear particles. Already much valuable information has been obtained about the lighter elements both by mass-spectrographic methods and by the study of atomic disintegrations. The discovery of the new radioactive elements has extended greatly the number of nuclei open to investigation, but since nearly all of these disintegrate by emitting either positrons or electrons forming a continuous spectrum we meet here the same difficulty in determining the total energy change in the disintegration as with the natural β-ray bodies. In this latter case Henderson* has proved the correctness of the suggestion of Ellis and Mott† that the difference of energy of two nuclei, apart from γ-emission, is given by the upper limit of the β-ray spectrum. However, as was emphasized by Cockroft at the British Association Meeting at Norwich, in September, 1935, this is a point which needs verification in the region of low atomic numbers and particularly for positron disintegration.

We have attempted to obtain some information on this point by investigating the disintegration of radio-phosphorus [P^{30}] formed from aluminium by α-particle bombardment. The disintegration of radio phosphorus has already been investigated several times, but there is such a notable disagreement between the values given by different observers for the energy of the upper limit that we felt fresh experiments were needed. Further, it is necessary to determine whether the upper limit corresponds to the formation of the ground state or of an excited state of the product nucleus.

Our final results are that the energy of the upper limit of the positrons from radio phosphorus is close to $2 \cdot 9 \times 10^6$ volts and that this corresponds to the formation of the product nucleus Si^{30} in its ground state.

---

* 'Proc. Roy. Soc.,' A, vol. 147, p. 572 (1934).

† 'Proc. Roy. Soc.,' A, vol. 141, p. 502 (1933).

Adopting the hypothesis that this energy determines the difference of internal energy of Si^{30} and P^{30} we find, on the basis of existing information, that there is the same difference of energy between Al^{27} and Si^{30}, whether Si^{30} is formed directly

$$_{13}Al^{27} + {}_2He^4 \rightarrow {}_{14}Si^{30} + {}_1H^1$$

or by the intermediate stage of P^{30}

$$_{13}Al^{27} + {}_2He^4 \rightarrow {}_{15}P^{30} + {}_0n'$$

$$_{15}P^{30} \rightarrow {}_{14}Si^{30} + \overset{+}{\varepsilon}.$$

Finally, if the mass of Si^{30} be written as $30 + x$, then the masses of Al^{27} and P^{30} are respectively $27 + x + 0{\cdot}0069$ and $30 + x + 0{\cdot}0042$.

2—PREVIOUS WORK

In describing the previous work on radio-phosphorus we shall also refer to data on the low energy portion of the spectrum, since in our former experiments we thought we obtained evidence of a definite lower limit to the curve.

Curie and Joliot* found that most of the positrons from radio-phosphorus were stopped by 1 gm/cm² of copper, suggesting an upper limit slightly greater than $2{\cdot}4 \times 10^6$ volts. This was confirmed approximately by cloud chamber measurements which gave a value of $3{\cdot}0 \times 10^6$ volts. No information was obtained about the form of the spectrum below 200,000 volts.

The present authors† investigated the absorption of the positrons in aluminium and found an end point at about 1·25 gm/cm² corresponding to an upper limit of $2{\cdot}75 \times 10^6$ volts. A rough magnetic analysis of the spectrum was also made which gave a value of about $2{\cdot}8 \times 10^6$ volts for the upper limit. A feature of the absorption curves found with both aluminium and copper, was a flat initial portion, and this was taken to indicate the absence of any very slow positrons. It was suggested that the distribution curve had a lower limit at about 200,000 volts.

Meitner‡ used a cloud chamber in a magnetic field to investigate the emission, and concluded that the upper limit lay between 2 and $2{\cdot}4 \times 10^6$ volts. While in some experiments she found a relatively large number of short tracks she showed that many of these must have been due to the γ-rays of the polonium used to form the radio-phosphorus and which, in some experiments, was kept in position during the expansion. However, after allowing for this effect, Meitner left it an open question whether the spectrum continued to zero energy or not.

---

* 'J. Phys. Rad.,' vol. 5, p. 153 (1934).

† 'Proc. Roy. Soc.,' A, vol. 146, p. 206 (1934).

‡ 'Naturwiss.,' vol. 22, p. 388 (1934).

Frisch[*] measured the absorption curve of the positrons in copper and did not observe any initial flat portion of the curve. He found a maximum range for the positrons of 0·8 gm/cm^2 corresponding to an upper limit of about 2×10^6 volts.

Nishina, Sagane, Takenchi, and Tomita[†] give a value for the upper limit of 4×10^6 volts based on the measurement of 1500 tracks in an expansion chamber. Only 10 of the tracks had energies greater than 3×10^6 volts. They observed positrons with energies as low as 100,000 volts present in appreciable numbers, but their experimental arrangement was not suitable for investigating the lower end of the spectrum.

Alichanow, Alichanian, and Dzelepow[‡] analysed the spectrum by the semi-circular magnetic focussing method and obtained a value $3 \cdot 75 \times 10^6$ volts for the upper limit. The low energy part of the spectrum was not investigated.

The lack of agreement between these various measurements of the upper limit may be seen from Table I; clearly further measurements are necessary.

TABLE I

Observer	Upper limit of the positron spectrum of P^{30} in volts
Curie and Joliot	3×10^6
Meitner	$2\text{–}2 \cdot 4 \times 10^6$
Frisch	2×10^6
Nishina *et alii*	4×10^6
Alichanow *et alii*	$3 \cdot 7 \times 10^6$
Ellis and Henderson	$2 \cdot 8 \times 10^6$

3—EXPERIMENTAL ARRANGEMENT

The arrangement of Geiger-Mueller counters used to investigate the absorption curve is shown drawn to scale in fig. 1. Only coincidences were counted using the circuit recently described by Barrasch.[§] It was hoped initially to investigate the absorption by inserting absorption sheets between the counters at B, but using the shape of source most suitable for activation, a small cylinder of aluminium foil, it proved difficult to canalize the beam sufficiently well to make the numbers of particles in the front and back counters comparable. As will be seen later, however, useful information was obtained by placing absorbers between the counters when there were initially absorbers in front of the first counter. The advantage of the arrangement depended really on the type of source available. The most powerful and convenient source of α-particles was a glass tube, containing radon, whose walls were sufficiently thin to allow the particles to escape. To use fully such an " α-ray tube " it was necessary to place over it an aluminium tube about 2 mm in diameter and about 2 to 3 cm long. Now any single counter which was of sufficient size to receive a reasonable number of the particles emitted by such a source would have a very large natural effect, whereas with our arrangement, as can be seen from fig. 1, we took in quite a large fraction of the particles but, owing to coincidence counting, the natural effect was small. Actually the natural effects of the front and back counters alone were 50 and 70 per minute respectively, whereas, using coincidence counting, they were only 5 per minute. Even this number was largely true coincidences of β-particles passing through both counters and due to the γ-rays from the laboratory supply of radium which was only 100 feet away. For example, an absorber of 0·7 gm/cm^2 placed between the counters reduced this natural count to 1·4 per minute.

---

[*] 'Nature,' vol. 133, p. 721 (1934).

[†] 'Sci. Pap. inst. phys. chem. Res. Tokyo,' vol. 25, p. 1 (1934).

[‡] 'Z. Physik,' vols. 5, 6, p. 350 (1934).

[§] 'Proc. Phys. Soc.,' vol. 47, p. 824 (1935).

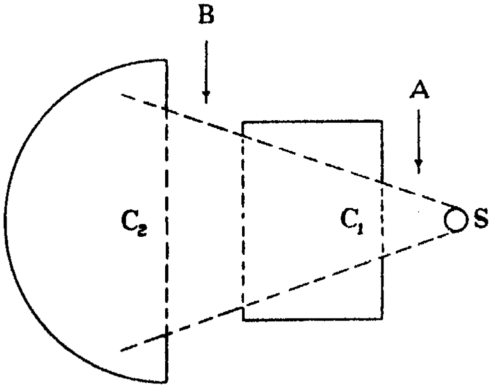


Fig. 1.

The radon tubes employed contained usually 60 to 100 millicuries of radon and their walls had a stopping power equivalent to 1·5 to 2·5 cm of air. The counters had sides, as shown, formed of brass grids covered with mica of about 2 cm air equivalent. The front counter had a cross-section $1 \cdot 9 \times 2 \cdot 7$ cm and the sensitive counting region was 7 cm long. The back counter was made from one-half of a brass tube of 4 cm internal diameter and had a counting region 10 cm long. The two counters were mounted 1 cm apart and the source of radio-phosphorus was placed at S, 1 cm from the face of the front counter.

4—EXPERIMENTAL RESULTS

The absorption curve was taken with copper, chiefly because it was less bulky than aluminium, and our measurements are shown in fig. 2. We should like to call attention particularly to the initial convex portion of the curve which extends up to about 0·09 gm/cm². With the β-ray bodies of the ordinary radioactive series no such initial convex portion is observed. This is the same result that we obtained in our previous experiments;* and it appears to indicate that there can be but few positrons emitted with less energy than about 200,000 volts.

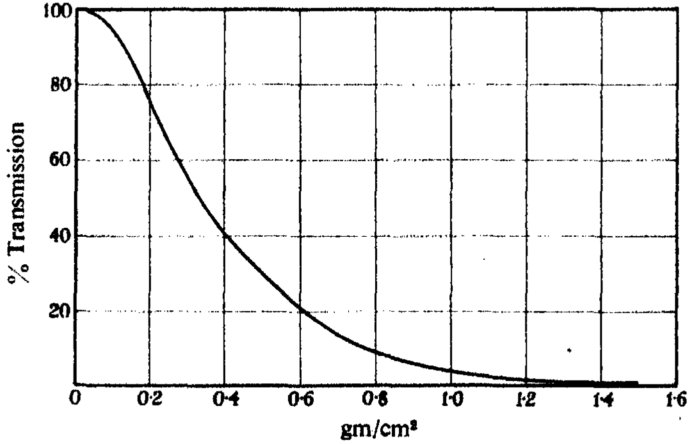


FIG. 2.

The curve as a whole agrees excellently with our previous measurements of the absorption in copper which were extended only to an absorption of 0·50 gm/cm². The end of the curve is shown on an enlarged scale in fig. 3, from which it will be seen that there is an end point at 1·30 ± 0·05 gm/cm².

While there is some doubt about the relation between the end point measured in copper in such an arrangement as this and the maximum energy to which it corresponds we can assign an upper energy limit to the main positron emission of $(2\cdot9 \pm 0\cdot1) \times 10^6$ volts.

---

* 'Proc. Roy. Soc.,' A, vol. 146, p. 256 (1934).

This assumes that the range-energy relation is the same for positrons as for electrons. This agrees with our previous measurements and with those of Curie and Joliot.*

We cannot at once, however, assume that this energy is really the energy of disintegration of radio-phosphorus [P^{30}] into silicon [Si^{30}]. From the work of Duncanson and Miller† we know that Si^{30} can be formed in at least three excited states of $2\cdot23$, $3\cdot60$, and $4\cdot74 \times 10^6$ volts excess energy, and the upper limit we have measured might correspond to any of these states of excitation, leading to a correspondingly greater energy of disintegration. However, we can be sure that this upper limit corresponds to the most frequent mode of disintegration, so that if, for example, it involved excitation to the $2\cdot23 \times 10^6$ volt state of Si^{30} we should find a γ-ray emission of $2\cdot23 \times 10^6$ volts for practically every disintegration. There is, in fact, as will be seen from fig. 3, a noticeable γ-ray effect after the end point at $1\cdot30$ gm/cm² and we investigated carefully whether this was simply the annihilation radiation of the positrons or whether there

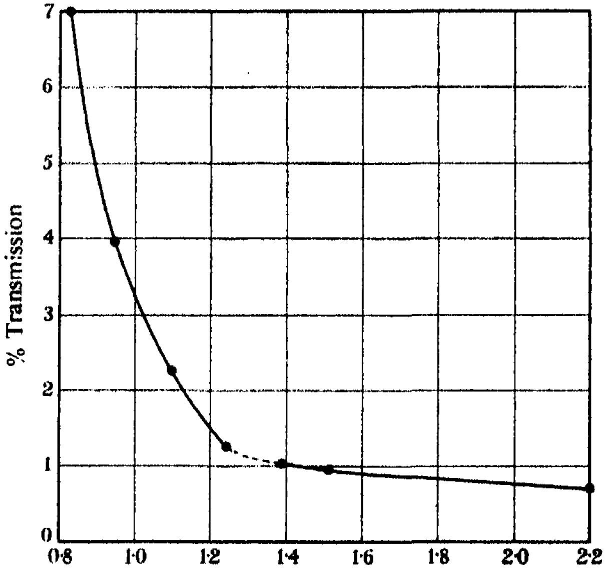


FIG. 3.

---

* 'J. Phys. Rad.,' vol. 5, p. 153 (1934).

† 'Proc. Roy. Soc.,' A, vol. 146, p. 396 (1934).

was in addition a γ-ray emission of about $2 \cdot 23 \times 10^6$ volts per disintegration.

The following two tests were made on the hardness of the radiation beyond the end point. Extra absorbing sheets of lead up to 5 gm/cm^2 were inserted between the source and the front counter and a mass absorption coefficient of about 0·2 was obtained. This is in fair agreement with other determinations of the absorption of the annihilation radiation of positrons, viz., Thibaud* 0·19, Joliot† 0·19–0·30, Crane and Lauritsen‡ 0·13, and Klemperer§ 0·15.

Using a source of Th C″ under exactly the same conditions, we obtained a mass absorption coefficient of about 0·095. Now per disintegration Th C″ emits 1 quantum of $2 \cdot 62 \times 10^6$ volts, 0·73 quantum of $0 \cdot 582 \times 10^6$ volts, 0·20 quantum of $0 \cdot 510 \times 10^6$ volts, and 0·08 quantum of $0 \cdot 275 \times 10^6$ volts.‖ We should not expect this radiation to be so markedly harder than the radio-phosphorus radiation if the latter really consisted of 2 quanta of $0 \cdot 511 \times 10^6$ volts (annihilation radiation) and also 1 quantum of $2 \cdot 23 \times 10^6$ volts. It would appear, therefore, that excitation of Si^{30} is not frequent.

In another experiment the source was placed inside a closed lead cylinder whose walls had a thickness of 2·05 gm/cm^2. The coincidences of the two counters are due to electrons ejected from this lead cylinder by the γ-rays. An estimate of the energy of these electrons, which are heterogeneous since they originate at various depths in the lead, was obtained by inserting absorbing sheets between the counters. It was found that 0·1 gm/cm^2 of copper produced a diminution of 20% which, for homogeneous rays, would indicate an energy of about 400,000 volts. The number of counts was too small to permit of a more detailed analysis.

It will be seen that both experiments suggest that the γ-radiation is mainly the annihilation radiation of the positrons and there is no indication of any large amount of hard radiation of over 2×10^6 volts energy.

In view, however, of the importance of this point, we attacked the problem in another way. We have already mentioned that the question is whether radio-phosphorus emits per disintegration only 2 quanta of $0 \cdot 511 \times 10^6$ volts or these two quanta and one quantum of $2 \cdot 23 \times 10^6$

---

* 'C.R. Acad. Sci., Paris,' vol. 197, p. 1629 (1933).

† 'C.R. Acad. Sci., Paris,' vol. 197, p. 1622 (1933); vol. 198, p. 1 (1934).

‡ 'Phys. Rev.,' vol. 45, p. 430 (1934).

§ 'Proc. Camb. Phil. Soc.,' vol. 30, p. 347 (1934).

‖ These intensities are slightly different from those given recently by Ellis ('Int. Conf. Nuclear Phys. London' (1934)), and are based on some unpublished work of F. Oppenheimer.

volts. In this latter case the comparison with the γ-radiation of Th C″ should be fairly close. We therefore attempted to compare the γ-radiation of these two bodies in a manner that would determine the relative number of quanta emitted. We prepared by recoil a source of Th C″ on the inside of an aluminium cylinder identical with those used for the production of radio-phosphorus. With this source we took an absorption curve in copper similar to that shown in fig. 2. The count due to γ-rays, shortly after the end point, was 10% of the initial intensity, whereas from radio-phosphorus the γ-rays were only about 1%.

This difference is decisive and we can conclude that the end point of the positron spectrum at $2 \cdot 9 \times 10^6$ volts corresponds to the formation of silicon in its ground state.

It may seem a little surprising that the difference between the γ-ray intensities of Th C″ and P^{30} was so great. This was due to two causes. Our experimental arrangement slightly favoured high-frequency γ-rays, since we only counted those electrons ejected from the absorbing sheets which passed through both counters. Secondly, the annihilation radiation from P^{30} is produced mainly in the copper absorbing sheets at the end of the range and represents, therefore, a diffuse source in contrast to the Th C″ γ-radiation which comes from the small activated aluminium tube.

It is interesting to recall that Duncanson and Miller* found that in the proton emission leading directly to Si^{30} the $2 \cdot 23 \times 10^6$ volt excited state is formed three times as often as the ground state. We must conclude that there is a difference in the relative frequency of formation of these two states by a factor of the order of 10 for the two reactions $Al^{27} + He^4 \rightarrow Si^{30} + H^1$ and $P^{30} \rightarrow Si^{30} + \overset{+}{\epsilon}$. It may be noted that when excitation does occur in the latter reaction it will give rise to a continuous β-ray spectrum ending at $0 \cdot 7 \times 10^6$ volts.

5—DISCUSSION

As suggested already by Meitner and Jaeckel,† we can obtain an important check on our application of energy relations to nuclear disintegrations by considering the dual disintegrations of aluminium under the action of α-particles. This may be written as follows. The equations will also represent the energy balance if the symbols are taken to represent the masses of the neutral atoms, electrons, and positrons, etc., and the

---

* 'Proc. Roy. Soc.,' A, vol. 146, p. 396 (1934).

† 'Z. Physik,' vol. 91, p. 493 (1934).

Q's represent in mass units the algebraic sum of the kinetic energies of the particles

$$ {}_{13}Al^{27} + {}_{2}He^{4} = {}_{14}Si^{30} + {}_{1}H^{1} + Q_1 \qquad (1)$$

$$ {}_{13}Al^{27} + {}_{2}He^{4} = {}_{15}P^{30} + {}_{0}n^{1} + Q_2 \qquad (2)$$

$$ {}_{15}P^{30} = {}_{14}Si^{30} + \overset{-}{\varepsilon} + \overset{+}{\varepsilon} + Q_3, \qquad (3)$$

whence

$$ {}_{1}H^{1} + Q_1 = \overset{-}{\varepsilon} + \overset{+}{\varepsilon} + {}_{0}n^{1} + Q_2 + Q_3. \qquad (4)$$

From Duncanson and Miller's* work on the production of protons from aluminium we obtain a value of Q_1 of $0 \cdot 0022_2$. Jaeckel's measurement on the maximum energy of the neutrons ejected from aluminium gives for Q_2 a value of $-0 \cdot 0022$. Our measurement of the P^{30} disintegration gives $Q_3 = 0 \cdot 0031$.

Hence

$$ {}_{0}n^{1} - {}_{1}H^{1} = 0 \cdot 0002_4. $$

This agrees excellently with the masses published recently by Oliphant, Kempton, and Rutherford,† viz., ${}_{0}n^{1} = 1 \cdot 0084$, ${}_{1}H^{1} = 1 \cdot 0081$. However, Aston's‡ new preliminary measurement of deuterium 2·0148, in conjunction with Chadwick and Goldhaber's§ measurement of the binding energy of the deuteron yields ${}_{0}n^{1} = 1 \cdot 009$. If this value were correct it would mean that there was roughly one million volts too much energy emitted in the neutron and positron branch. The measurement of neutron energies is admittedly difficult, but as Jaeckel pointed out his value was more likely to be too small than too large. Our measurement of the energy of disintegration is shown with an uncertainty of 100,000 volts which is probably an over-estimate of the error. It would tend to be too low from the inherent characteristics of the range method, but, on the other hand, we have made a generous allowance for the difference between the range-energy curves in aluminium and copper, which, if anything, has put the value too high. Provisionally it seems reasonable to accept the " disintegration masses " for the neutron-proton difference and hence conclude that it is correct to deduce the energy of a positron disintegration from the energy of the upper limit of the spectrum.

It may be convenient to state here the relation masses of Al^{27}, P^{30}, and Si^{30}. If the, at present, unknown mass of Si^{30} be written as $30 + x$ then $Al^{27} = 27 + x + 0 \cdot 0069$ and $P^{30} = 30 + x + 0 \cdot 0042$.

---

* 'Proc. Roy. Soc.,' A, vol. 146, p. 396 (1934).

† 'Proc. Roy. Soc.,' A, vol. 150, p. 241 (1935).

‡ 'Nature,' vol. 135, p. 541 (1935).

§ 'Proc. Roy. Soc.,' A, vol. 151, p. 479 (1935).

In conclusion, we should like to thank Mr. G. R. Crowe for preparing the radon tubes, and Mr. R. Cole for help in the experiments.

SUMMARY

The maximum energy of the positrons emitted from radio-phosphorus P^{30} has been determined by measuring the end point of the absorption curve in copper. The value found is $2 \cdot 9 \pm 0 \cdot 1 \times 10^6$ volts in good agreement with a previous determination. A careful search was made to see if any γ-rays accompanied the disintegration but none was found besides the annihilation radiation of the positrons. This energy of $2 \cdot 9 \times 10^6$ volts therefore determines the difference of energy of the ground states of P^{30} and Si^{30}. Combining this value with other published data, the difference in mass of Al^{27} and Si^{30} is found to be the same when calculated from

$${}_{13}Al^{27} + {}_{2}He^{4} \rightarrow {}_{14}Si^{30} + {}_{1}H^{1},$$

or from the two stages

$${}_{13}Al^{27} + {}_{2}He^{4} \rightarrow {}_{15}P^{30} + {}_{0}n^{1}$$

$${}_{15}P^{30} \rightarrow {}_{14}Si^{30} + \overset{+}{\varepsilon}.$$

INDEX TO VOLUME CLII (A)

Alloys, effect of thermal agitation on atomic arrangements, III (Williams), 231.
Ammonia and deutero-ammonia, decomposition, effect of imprisonment of resonance radiation (Melville), 325.
Antimony, explosive, II (Coffin), 47.
Atomic arrangements, effect of thermal agitation in alloys, III (Williams), 231.

Bhabha (H. J.) The creation of electron pairs by fast charged particles, 559.
Brindley (G. W.) and Hoare (F. E.) The diamagnetic susceptibilities of salts forming ions with inert gas configurations. II—The alkaline halides, 342.

Coffin (C. C.) Studies on explosive antimony. II—Its structure, electrical conductivity and rate of crystallization, 47.
Collens (H.) *See* Schonland and others.
Combustion, influence of molecular structure on (Ubbelohde), 354.
Cooper (E. R.) *See* Darbyshire and Cooper.
Cunnold (F. A.) The optical system of the disappearing filament pyrometer, 64.
Current system, aspects producing magnetic disturbance (Stagg), 277.

Darbyshire (J. A.) and Cooper (E. R.) Diffraction of electrons by metal crystals and by mica, 104.
Deuterium, ortho-, thermal conversion (Farkas and Farkas), 124.
Deuteron, ratio of magnetic moment (Farkas and Farkas), 152.
Diamagnetism, susceptibilities of salts forming ions with inert gas configurations, II (Brindley and Hoare), 342.
Diffraction of electrons by metal crystals and by mica (Darbyshire and Cooper), 104.
Dipole induction and the solvent effect in dipole moment measurements (Frank), 171.
Discussion on supraconductivity and other low temperature phenomena, 1.

Eddington (Sir Arthur) The pressure of a degenerate electron gas and related problems, 253.
Electric discharge, chemical reaction (Willey), 158.
Electrolytes, strong, statistical treatment (Levine), 529.
Electron, free, temperature dependence of susceptibility (Stoner), 672.
Electron gas, degenerate, pressure and related problems (Eddington), 253.
Electron pairs, creation by fast charged particles (Bhabha), 559.
Electrons, diffraction by metal crystals and by mica (Darbyshire and Cooper), 104.
Ellis (C. D.) and Henderson (W. J.) The energy of disintegration of radio-phosphorus [P^{30}], 714.
Energy exchange between diatomic gas molecules (Jackson and Howarth), 515.
Equilibrium-phase of supraconductors in a magnetic field (London), 650.
Explosions, sensitized, I (Foord and Norrish), 196.

Farkas (A.) and Farkas (L.) Experiments on heavy hydrogen. V—The elementary reactions of light and heavy hydrogen. The thermal conversion of orthodeuterium and the interaction of hydrogen and deuterium, 124.

Farkas (A.) and Farkas (L.) Experiments on heavy hydrogen. VI—The ratio of the magnetic moment of the proton to the magnetic moment of the deuteron, 152.
Farkas (L.) *See* Farkas and Farkas.
Field, relativistic self-consistent (Swirles), 625.
Films, adsorbed, of oxygen on tungsten (Roberts), 464.
Films, composite, of oxygen and hydrogen on tungsten (Roberts), 477.
Fluid, instability when heated from below (Schmidt and Milverton), 586.
Foord (S. G.) and Norrish (R. G. W.) A study of sensitized explosions. I—The hydrogen oxygen reaction catalysed by nitrogen peroxide, 196.
Frank (F. C.) Dipole induction and the solvent effect in dipole moment measurements, 171.

Gases, diffusion through metals, II (Smithells and Ransley), 706.
Goodeve (C. F.) and Taylor (A. W. C.) The continuous absorption spectrum of hydrogen bromide, 221.

Halides, alkaline (Brindley and Hoare), 342.
Henderson (W. J.) *See* Ellis and Henderson.
Hoare (F. E.) *See* Brindley and Hoare.
Howarth (A.) *See* Jackson and Howarth.
Hydrocarbons, aromatic, ultra-violet absorption spectra, I (Mayneord and Roe), 299.
Hydrocarbons, investigations of combustion, I and II (Ubbelohde), 354, 378.
Hydrogen, absorption on tungsten (Roberts), 445.
Hydrogen bromide, continuous absorption spectrum (Goodeve and Taylor), 221.
Hydrogen, ground state of molecule (Richardson), 503.
Hydrogen, heavy, V and VI (Farkas and Farkas), 124, 152.

Iodine, nuclear spin, II (Tolansky), 663.

Jackson (J. M.) and Howarth (A.) Exchange of energy between diatomic gas molecules and a solid surface, 515.

Levine (S.) Statistical treatment of strong electrolytes, 529.
Lightning, progressive (Schonland and others), 595.
London (H.) Phase-equilibrium of supraconductors in a magnetic field, 650.

Magnetic disturbance, produced by aspects of current system (Stagg), 277.
Malan (D. J.) *See* Schonland and others.
Massey (H. S. W.) and Mohr (C. B. O.) The interaction of light nuclei. II—The binding energies of the nuclei H_1^3 and He_2^3, 693.
Mayneord (W. V.) and Roe (E. M. F.) The ultra-violet absorption spectra of some complex aromatic hydrocarbons—I, 299.
Melville (H. W.) The effect of imprisonment of resonance radiation in the decomposition of ammonia and of deutero-ammonia, 325.
Metallic systems, binary (Stockdale), 81.
Metals, diffusion of gases through, II (Smithells and Ransley), 706.
Milverton (S. W.) *See* Schmidt and Milverton.
Mohr (C. B. O.) *See* Massey and Mohr.

Molecules, diatomic, gas, exchange of energy with a solid surface (Jackson and Howarth), 515.
Møller (C.)—On the radiative collision between fast charged particles, 481.

Nuclear reactions, energies (Wilson), 497.
Nuclei, light, interaction, II (Massey and Mohr), 693.

Oxygen, preparation by electrolysis of baryta (Stoddart), 273.

Particles, fast charged, radiative collision (Møller), 481.
Payman (W.) and Titman (H.) Explosion waves and shock waves. III—The initiation of detonation in mixtures of ethylene and oxygen and of carbon monoxide and oxygen, 418.
Phosphorus, radio, energy of disintegration (Ellis and Henderson), 714.
Proton, magnetic moment (Farkas and Farkas), 152.
Pyrometer, disappearing filament, optical system (Cunnold), 64.

Radiation, resonance, effect of imprisonment in the decomposition of ammonia and of deutero-ammonia (Melville), 325.
Radiative collision between fast charged particles (Møller), 481.
Ransley (C. E.) *See* Smithells and Ransley.
Read (J.) The absorption of monochromatic X-ray beams, of wave-length in the region 50 to 20 *x*-units, in lead, tin, copper, and iron, 402.
Richardson (O. W.) On the ground state of (H_2), the molecular ion (H_2+) and wave mechanics, 503.
Roberts (J. K.) Composite films of oxygen and hydrogen on tungsten, 477.
Roberts (J. K.) Some properties of adsorbed films of oxygen on tungsten, 464.
Roberts (J. K.) The adsorption of hydrogen on tungsten, 445.
Roe (E. M. F.) *See* Mayneord and Roe.

Schmidt (R. J.) and Milverton (S. W.) On the instability of a fluid when heated from below, 586.
Schonland (B. F. J.), Malan (D. J.), and Collens (H.) Progressive lightning, II, 595.
Smithells (C. J.) and Ransley (C. E.) The diffusion of gases through metals. II—Diffusion of hydrogen through aluminium, 706.
Spectra absorption of hydrocarbon combustion (Ubbelohde), 378.
Spectra, ultra-violet absorption, of some complex aromatic hydrocarbons, I (Mayneord and Roe), 299.
Spectrum, arc, and fine structure of iodine (Tolansky), 663.
Spectrum, continuous absorption, of hydrogen bromide (Goodeve and Taylor), 221.
Stagg (J. M.) Aspects of the current system producing magnetic disturbance, 277.
Stockdale (D.) Numerical relationships in binary metallic systems, 81.
Stoddart (E. M.) The preparation of oxygen by electrolysis of baryta, 273.
Stoner (E. C.) The temperature dependence of free electron susceptibility, 672.
Supraconductivity, discussion, 1.
Supraconductors in a magnetic field, phase-equilibrium (London), 650.
Surface, solid, exchange of energy between diatomic gas molecules (Jackson and Howarth), 515.
Swirles (Bertha) The relativistic self-consistent field, 625.

Taylor (A. W. C.) *See* Goodeve and Taylor.

Titman (H.) *See* Payman and Titman.

Tolansky (S.) The nuclear spin of iodine. II—Fine structure in the arc spectrum and a fine structure perturbation effect, 663.

Tungsten, adsorption of hydrogen (Roberts), 445.

Ubbelohde (A. R.) Investigations on the combustion of hydrocarbons. I—The influence [illegible] molecular structure on hydrocarbon combustion; II—Absorption spectra and chemical properties of intermediates, 354, 378.

Wave mechanics and molecular hydrogen (Richardson), 503.

Waves, explosion and shock, III (Payman and Titman), 418.

Willey (E. J. B.) On chemical reaction in the electric discharge. I—The chemical effects of impulse discharges, 158.

Williams (E. J.) Effect of thermal agitation on atomic arrangements in alloys, III, 231.

Wilson (H. A.) The energies of nuclear reactions, 497.

X-ray beams, monochromatic, absorption in lead, tin, copper, and iron (Read), 402.

END OF THE ONE HUNDRED AND FIFTY-SECOND VOLUME (SERIES A)

I. A. R. I. 75.

INDIAN AGRICULTURAL RESEARCH
INSTITUTE LIBRARY,
NEW DELHI.

Date of issue	Date of issue	Date of issue
3.12.58		

GIPN—S4—34 I. A. R. I. 56.—22-4-57—4,000.